The Genomics of Adaptation to Climate across Latitude and Elevation in the European Common Frog

---

Dissertation

zur

Erlangung der naturwissenschaftlichen Doktorwürde
(Dr. sc. nat.)

vorgelegt der

Mathematisch-naturwissenschaftlichen Fakultät

der

Universität Zürich

von

Alexandra Jansen van Rensburg

aus

Südafrika

Promotionskommission

PD Dr. Josh Van Buskirk (Leitung der Dissertation)

Prof. Dr. Lukas Keller (Vorsitz)

Prof. Dr. Rolf Holderegger

Dr. Daniel Berner

Zürich, 2018

The Genomics of Adaptation to Climate Across Latitude and Elevation in the European Common Frog

PhD Thesis

Alexandra Jansen van Rensburg

Institute of Evolutionary Biology and Environmental Studies
University of Zurich, Switzerland

Universität Zürich[UZH]

Summary

In this thesis I investigate how demographic history and environment determine the geographic distribution of genomic variation in the European common frog. I show that colonisation history and gene flow are important determinants of population structure, but that adaptation to environment underlies a large proportion of genomic variation across the species range.

Rapidly changing climate in the last century has had a profound effect on species distributions. Research into the rapid response of species to these changes have shown that the geographic distribution of genetic variation across a species range is an important determinant of population persistence. The geographic distribution of genetic variation is determined by colonisation history and gene-flow between populations, and adaptive divergence across environmental gradients.

In this thesis I investigate the genomics of adaptation to environment in the European common frog. *R. temporaria* is widespread across Europe and occurs across a diverse range of habitats. Similar evironmental gradients occur across latitude in Scandinavia and elevation in the European Alps, and there is evidence for adaptive genetic divergence across both gradients. This provides an ideal system to investigate the genomics of adaptation to similar environmental gradients under different selection-migration regimes.

First, I investigate how colonisation history has determined the distribution of genetic variation in Switzerland in a phylogeographic study of *R. temporaria* in the Swiss Alps. I find two mitochondrial lineages distributed roughly north and south of the alpine ridge bisecting Switzerland with a contact zone in eastern Switzerland. I show that these lineages likely originate from two different refugia in northern Italy, and colonised Switzerland via trans- and circum-alpine routes.

Next, I establish genome-wide molecular markers to investigate the distribution of neutral genetic variation across Switzerland. I find that

population structure reflects their colonisation history and identify four geographic lineages.

Third, I investigate how much genetic variation can be explained by geographic distance and climate within each geographic region identified in Chapter 2. I find that geographic distance and climate are important at different geographic scales, but that climate explained a large proportion of genomic variation. I also show that different climatic variables are important to populations in northern and southern Switzerland.

Finally, I investigate how population structure and climate uderlie genomic variation across the latitudinal gradient in Scandinavia. I also show that climate explains a large proportion of genomic variation despite strong population structure across the gradient.

The work presented here is an important contribution towards understanding the environmental drivers and geographic scale of adaptive divergence in this widespread amphibian.

Zusammenfassung

In dieser Arbeit untersuche ich, in welchem Ausmass die demographische Geschichte und die Umwelt die geographische Verbreitung der genomischen Variation des Europäischen Grasfrosches bestimmt. Ich zeige auf, dass die Geschichte der Kolonisierung und der Genfluss wichtige Faktoren für die Erkärung der Populationsstruktur darstellen, dass aber ein grosser Teil der genomischen Variation über das Verbreitunggebiet der Art durch die Anpassung an die Umwelt erklärt werden kann.

Das sich rasch verändernde Klima im letzten Jahrhundert hatte einen tiefgreifenden Einfluss auf die Verbeitung von Arten. Untersuchungen über die rasche Anpassung von Arten an diese Veränderungen haben gezeigt, dass die geographische Verbreitung der genetischen Variation über das Verbreiuntgsgebiet einer Art ein wichtiger Faktor für den Fortbestand einer Art darstellt. Die geographische Verbreitung der genetischen Variation wird einerseits durch die Geschichte der Kolonisierung und des Genflusses zwischen Populationen und andereseits durch das adaptive Auseinanderdriften entlang von Umweltgradienten bestimmt.

In dieser Arbeit untersuche ich die Genomik der Adaptation des Europäischen Grasfrosches an die Umwelt. Rana temporaria ist in Europa weit verbreitet und kann in verschiedensten Habitaten existieren. Es bestehen umfassende Hinweise auf eine adaptive genetische Aufspaltung entlang des Nord-Südgradienten in Skandinavien. Ein ähnlicher Umweltgradient existiert ebenfalls entlang von verschiedenen Höhenstufen in den Europäischen Alpen, in welchen R. temporaria trotz eines hohen Grads an Genfluss adaptive Aufspaltung aufweist. Dies stellt ein ideales System für die Untersuchung der Genomik der Adaptation an ähnliche Umweltgradienten unter verschiedenen Selektions- Migrationsregimes dar.

Zu Beginn untersuche ich mittels einer phylogeographischen Studie in den Schweizer Alpen, wie die Geschichte der Kolonisierung die Verbreitung der genetischen Variation von R. temporaria in der Schweiz bestimmt hat. Ich

finde zwei mitochondriale Abstammungslinien welche nördlich und südlich des Alpenkamms verlaufen, der die Schweiz unterteilt, mit einer Kontaktzone in der Ostschweiz. Ich zeige auf, dass diese Abstammungslinien vermutlich von zwei unterschiedlichen Rückzugsgebieten in Norditalien ausgingen, und die Scheiz via trans und circum-alpine Routen kolonisierten.

Im nächsten Schritt, etabliere ich genomweite molekulare Marker, um die Verbreitung von neutraler geneticher Variation in der Schweiz zu untersuchen. Ich zeige auf, dass die Populationsstruktur die Geschichte der Koloniesierung widerspiegelt und identifiziere vier geographische Abstammungslinien.

Drittens untersuche ich, wieviel der genetischen Variation durch die geographische Distanz und das Klima innerhalb der vier geographischen Regionen, welche im zweiten Kapitel gefunden wurden, erklärt werden kann. Ich zeige auf, dass die geographische Distanz und das Klima auf unterschiedlichen geographischen Skalen eine wichtige Rolle spielen, dass aber das Klima einen Grossteil der genomischen Variation erklärt. Darüberhinaus zeige ich, dass für Populationen in der Nord- und Südschweiz unterschiedliche klimatische Variablen von Bedeutung sind.

Zuletzt untersuche ich, wie die Populationsstruktur und das Klima der genomischen Variation entlang von Nord-Südgradienten in Skandinavien zugrunde liegen. Ausserdem zeige ich, dass das Klima, trotz starker Populationsstruktur, einen grossen Teil der genomischen Variation erklärt und identifiziere einen adaptiven Temperatur-Schwellenwert in Südschweden.

Die hier vorgestellte Arbeit leistet einen wichtigen Beitrag zum Verständnis der Umweltfaktoren und der geographischen Skalen, welche eine adaptiven Aufspaltung dieser weitverbreiteten Amphibienart hervorrufen.

Acknowledegements

The last five years have been enormously stimulating and challenging, and I am grateful to my friends, family, peers, and supervisors for their support and contributions during this time.

The person due the most thanks is **Josh Van Buskirk**. Thank you for giving me the opportunity to come to Switzerland, and work on this extremely interesting project. Josh has given me enormous freedom to pursue all my ideas (I grudgingly admit some were ridiculous). He's been very patient and supportive as I struggled my way through the lab work and bioinformatics. Thank you for always being available, particularly during the last year of writing and analysis! I'm especially grateful for all the critical and in-depth discussions and comments on my work. I learned so much from you!

My thesis committee, **Lukas Keller, Rolf Holderegger,** and **Daniel Berner**, provided invaluable support and advice during the every committee meeting**.** Thanks to **Daniel** for advice on the genomic library preparation, and for letting me observe your lab work. **Rolf** was very generous with his time, and gave me a lot of good advice, particularly at the start of the PhD. I'm very grateful to **Lukas** who welcomed me into his group and always made time for me when I needed advice.

The **Keller, Postma,** and **Van Buskirk** groups provided a very stimulating and entertaining environment. Thanks to everyone for all the coffee breaks, meetings, chats, and after work beers: Andres, Barbara, Beni, Beatrice, Christine, Debbie, Dominique, Erica, Erik, Eva, Franzisca, Glauco, Kasia, Irene, Jael, Judith, Lukas, Mathieu, Mirco, Nicolas, Nina, Philip, Pirmin, Rassim, Rien, Simon, Timotheé, Uli, Ursina, Vanja.

The huge sampling effort during my first year was only made possible with the help of several people. **Josh** who helped coordinate and plan the field work initially, accompanied me to the first few ponds, and taught me everything I needed to know about finding frog spawn. **Nicolas Pruvost** assisted in raising tadpoles while I was out in the field. **Tim Bray** was a massive help in building and gathering equipment, dip-netting, exploratory forrays to isolated ponds, navigating, and attempting to explain our presence

to curious passers-by. Thanks to **Mathieu Robin**, who generated a lot of the mitochondrial data and helped with Chapter 1.

All my data was generated at the Genomic Diversity Centre, ETHZ, Zurich. I spent almost 2 years here, preparing and optimising the ddRAD protocol, and generating all of the data. I'm very grateful to **Sylvia Skobel** and **Aria Minder** for being constantly available to answer questions or help me when needed. **Claudia Michel** was a great help when I first started the lab work. Her extensive experience with ddRAD lab work was an invaluable resource. Thanks also for bioinformatics support from **Stefan Zoller, Stefan Zoller,** and **Stefan Zoller.** I'm particularly grateful to **Stefan** whose help with the de novo assembly pipeline reduced the computation time massively. Thank you for taking so much time to help me.

The Functional Genomcs Center at UZH provided excellent sequencing facilities and bioinformatic support throuhgout the project. **Catherine Aquino** was particularly helpful and sequenced all of the libraries. Thank you very much for being available all the time and for working so hard to provide me with excellent data! I'm also very grateful to **Hubert Rehrauer** who has been extremely responsive and helpful with regard to bioinformatics questions or help installing software on the server. Finally, **Kentaro Shimizu** generously allowed me to store data and perform analyses on his groups server. I'm also grateful to **Masa Hatakeyama** who provided rapid and efficient bioinformatics support while I was using this server.

I'm very grateful to **Isabel Schöechli,** our institutional secretary, who was an invaluable resource in helping me navigate all the permits and paperwork I needed throughout. Similarly I greatly benefited from **Ursina Tobler's** help with contracts, reference letters, permits, insurance, and various other official documents. Thank you both! If it weren't for you I have no doubt I'd still be sitting in a queue somewhere with the wrong form completed for something.

I had the privilege of spending some time in **Victoria Sork's** lab at UCLA during my PhD. Victoria, thank you for welcoming me into your lab and into your home. You were incrediby supportive and helpful during my stay and after. I learned a great deal from you and from my interactions with the

members of your group. Thank you also for agreeing to be the external examinor of this thesis.

My big office at the end of the corridor not only got the most sun, but included a steady flow of interesting museum specimen. It also came with an amazing set of office mates. **Jasmin Winkler, Irene Weinberger, Christine Grossen,** and **Simon Evans,** thank you for making the office such an enjoyable environment to work in. I miss all the ridiculous, serious, funny, disgruntled, happy, and supportive interactions!

A big thanks to all the friends I haven't mentioned yet. **Sam Cruickshank**, **Reini Stockenhuber**, **Annina Brunner**, **Sofia Scherrer**, **Nina Vasiljevic**, **Benedict Gehr**, **Glauco Camenisch, Jobran Chebib** thanks for all the braais, games, coffees, and chats. I'm looking forward to many more! **Preya Subramoney, Emilie Boissin, Nienke Stols,** thank you for checking up on me, and for all the adventurous breaks. Thanks to **Alex Kolodotschko** who gave me a place to stay in Bristol. More importantly thank you for reminding me about life away from the computer; Mostly via the medium of food and drink, and the occasional zombie/pirate party.

A PhD acknowledgement section is one of the rare opportunities one has to formally and publicly acknowledge the huge role that family plays in your life. I am so grateful for all the support and encouragement that I've received throughout my schooling and higher education. I would certainly not be here without the financial support I enjoyed, but I am even more grateful for the emotional support and the unwavering confidence my parents have in me! **Mam**, **Dad**, baie dankie vir al julle ondersteuning.

'n Baie groot dankie ook aan die res van my familie: **Ouma Lucille, Oupa Bert, Oom Paul, Tannie Charmaine, Ingrid, Jonathan, Gustav, Ouma Rosa.** Om te sien hoe passievol, talentvol, hard-werkend, gemotiveerd, en avontuurlustig elkeen van julle is laat my voel asof daar geen limiet is aan wat ek kan bereik nie. Ek voel baie geëerd om deel van ons familie te wees.

I've also acquired a whole new family during my PhD. Thank you **John, Janet, Jon, & Sarah** for all the support and motivation throughout.

Finally, I would like to thank **Tim Bray**. Thank you for moving to Switzerland with me and for believing in me. I'm so happy that I got to share

the field work experience and exploring Switzerland with you. I'm looking forward to our post-PhD lives together!

Curriculum Vitae

Personal

Surname	**JANSEN VAN RENSBURG**
First names	**Alexandra**
Date of birth	**01 January 1986**
Nationality	**South Africa**

Education

01.2013 - 02.2018 **PhD in Evolutionary Biology at the University of Zurich**
Thesis title: *The Genomics of Adaptation to Climate Across Latitude and Elevation in the European Common Frog*

02.2008 - 06.2011 **MSc in Evolutionary Biology at the University of Pretoria**
Thesis title: *Resolving the fine-scale population structure of Nesospiza buntings using a molecular multi-markers system*

02.2007 - 01.2008 **BSc Genetics: Honours at the University of Pretoria**
Thesis title: *The inference of early human migration in Africa using genetic diversity within the gastric pathogen Helicobacer pylori*

02.2004 - 12.2006 **BSc Human Genetics at the University of Pretoria**

01.1999 - 12.2003 **Matriculation at St. Patrick's College, Kokstad**
Subjects: Science, Biology, Accountancy

Table of Contents

General Introduction

Anthropogenic climate change has had a profound effect on global biodiversity. Mismatches between species phenotypes and environment have resulted in shifts in species ranges, reduced population sizes, and even extinction (Parmesan 2006). A major challenge in mitigating further loss of biodiversity lies in understanding how species adapt to their environments, and predicting their ability to adapt to future environments. Recent studies have shown that the availability of genetic diversity is a good predictor of adaptive resilience.

Aided by the rapid increase in the number of genome wide markers that can be generated for non-model organisms, there has been a rapid development in methods to investigate adaptive response to environment in non-model organisms. Two main approaches are used to identify potentially adaptive loci: F_{ST} outlier methods and environmental association analyses (Vitti *et al.* 2013; Rellstab *et al.* 2015). These identify loci that are likely to be under divergent selection between populations based on F_{ST} distributions, or identify loci associated with specific environmental variables. Once these loci are identified, they can be used (for example) to test for parallelism in adaptatation between geographic regions or species, or to identify specific genes or gene networks involved in adaptation (Stapley *et al.* 2010). Furthermore, multivariate analyses have proven useful tools to examine the multivariate nature of adaptation to environment, where adaptive divergence at multiple loci is driven by multiple climate varaibles (Sork *et al.* 2013). Recently, the incorporation of machine learning methods such as gradient forest analyses have enabled the characterisation of non-linear changes in allele frequencies associated with climate variables, and to identify potential adaptive thresholds along these variables (Fitzpatrick & Keller 2015; Bay *et al.* 2018). Altogether genome-wide datasets in combination with the statistical developments in the field of adaptation genomics provides a powerful approach to study adaptation to heterogeneous environments.

Theoretical and empirical studies have determined some general expectations regarding adaptation to heterogeneous environments. One is

that adaptation often occurs through soft selection - that is selection on loci that are already in the population at low frequencies as opposed to new mutations (Hermisson & Pennings 2005; Pennings & Hermisson 2006). Secondly, the phenotypes that underlie adaptation to environment are most commonly polygenic and involve small changes in the allele frequency of multiple loci of small effect size (Pritchard & Di Rienzo 2010; Yeaman 2015). This implies that the starting allele frequencies in populations are important determinants of whether selection acts on them, and this could reduce the amount of parallelism (at the locus level) in adaptation across a species range (e.g. Rellstab *et al.* 2016).

Adaptation to the local environment depends on changes in allele frequencies at many loci, but also that these adaptive genotypes are maintained at a high frequency in the population. Gene-flow can counteract adaptation by reducing the frequency of the adaptive genotypes (swamping) or by breaking down adaptive genotypes through recombination (Lenormand 2002). In order to maintain adaptive divergence across an environmental gradient, directional selection has to be stronger than gene flow. This presents a particularly interesting case for studying adaptive divergence at different geographic scales.

Steep environmental gradients occur across a geographic distance smaller than the individual's dispersal distance. Elevation presents a steep environmental gradient for many species. Latitudinal gradients present good contrasts to elevation, since the same environmental distance is found at a much larger geographic scale. Hence the migration-selection balance required for adaptive divergence is very different.

Study system

In this thesis I focus on the European common frog, *R. temporaria*, which are found throughout Europe (Sillero *et al.* 2014). Amphibians are particularly interesting for the study of adaptive divergence given the relative ease with which many phenotypic traits can be measured (Miaud & Merilä 2001). *R. temporaria* occur across a wide range of habitats, and there is extensive evidence for adaptive divergence between different environments

(e.g. Laugen *et al.* 2003; Palo *et al.* 2003; Bachmann 2017). Here I study populations across their elevational distribution in the Swiss Alps and latitudinal distribution in Scandinavia. These similar environmental gradients occur at very different geographic scales. This presents an ideal system within which to examine adaptation to environment across very different gene-flow and selection gradients. The elevational gradient is approximately 2000 m, thus presents a very steep gradient. Common garden experiments have shown adaptive divergence across elevation in the Alps despite high levels of gene flow (Bachmann 2017). The latitudinal gradient, on the other hand, occurs across ~ 1500 km, thus represents a much more shallow selection gradient.

Outline of Thesis

The aim of this thesis was to understand how neutral and adaptive processes have shaped the geographic distribution of genomic variation in *R. temporaria* across elevation and latitude. The geographic distribution of genetic diversity is determined by a combination of colonisation history, gene-flow between populations, and adaptive divergence between populations. Theory suggests that genetic priority effects will determine genetic population structure; that is, the order in which genotypes colonise an environment will affect whether they are maintained in the population or not (Orsini *et al.* 2013; Meester *et al.* 2016). In combination with the prevalence of soft selection underlying adaptation to environment, this implies that understanding the colonisation history and neutral population structure of the study system is imperative when studying adaptation to environment. Therefore I first investigate how colonisation history and contemporary gene-flow have determined neutral population structure across Switzerland. This is described in the first two chapters. In the third and fourth chapters I explore the genomics of adaptation to environment across the elevational and latitudinal gradients.

In **Chapter 1** I investigate the phylogeographic history of *R. temporaria* in the Swiss Alps. I find two mitochondrial lineages distributed roughly north and south of the alpine ridge bisecting Switzerland with a contact zone in

eastern Switzerland. These lineages diverged at the onset of the last glacial cold period. I show that these lineages likely originate from two different refugia in northern Italy, and colonised Switzerland via trans- and circum-alpine routes. These results show that the *R. temporaria* colonisation history has resulted in two deeply diverged genetic lineages, and hence different pools of genetic variation, occuring in northern and southern Switzerland.

In **Chapter 2,** I establish genome-wide molecular markers to investigate the distribution of neutral genetic variation across Switzerland. I find that population structure reflects their colonisation history and identify four geographic lineages. I find that evidence for high gene flow across elevation, and that a large portion of genetic diversity can be explained by elevation. Together these results imply that adaptive divergence across the environmental gradient is maintained despite high levels of gene flow.

In **Chapter 3**, I use a multivariate regression analysis to investigate how much genetic variation can be explained by geographic distance and climate within each geographic region identified in Chapter 2. I find that geographic distance and climate are important at different geographic scales, but that climate explained a large proportion of genomic variation. I also show that different climatic variables are important to populations in northern and southern Switzerland; a measure of Spring snow melt in the north, and temperature and solar radiation in the south.

In **Chapter 4,** I investigate how population structure and climate have structured genomic variation across the latitudinal gradient in Scandinavia. I find that climate explains a large proportion of genomic variation despite strong population structure across the gradient. Climate and geographically structured climate account for more than half of the genomic variation across latitude. Next, I identify geographic regions where allele frequencies of adaptive loci change more rapidly than expected. I identify a threshold response to a temperature related variable which results in dramatically different allele frequencies in two regions along the gradient. Finally, I find that the change in allele frequency between neutral and adaptive loci are the most different in southern Sweden, suggesting that this region may be particularly susceptible to climate change.

In conclusion, neutral and adaptive processes have shaped the geographic distribution of genomic variation across elevation and latitude in *R. temopraria*. I find that the Alps were a semi-permeable barrier to colonisation for *R. temporaria,* and that both northern and southern Switzerland were likely colonised from multiple refugia in the Italian Alps (Chapter 1). This is a surprising result given that the Alps restricted many lineages in the Italian peninsula after the retreat of the glaciers (Hewitt 1999). Colonisation by multiple mitochondrial lineages and subsequent contact zones between them has resulted in unexpectedly high genetic diversity in the Switzerland.

Further, I find that population structure in the Alps reflects colonisation history, likely because the complex topography of the Alps reduces gene flow over large distances (Chapter 3). However, I find high levels of gene flow across elevation along with evidence for adaptive divergence across elevation. Adaptive divergence in several larval life history traits has been found across elevation in Switzerland based on common garden experiments (Bachmann 2017). The results from my thesis show that a large portion of genomic variation is likely associated with this adaptive divergence, and that climate describes more genomic variation at regional scales than geographic distance does (Chapter 4). In addition, I find regional differences in the importance of various climate variables. Specifically, a measure of snow melt, which is highly correlated with elevation, is important in northern Switzerland, while precipitation and solar radiation are important in southern Switzerland. When investigating the latitudinal gradient I also find a large proportion of genomic variation explained by climate independent of geography, despite the fact that colonisation history and population structure is largely confounded with geography.

In summary, this thesis shows that a large proportion of genomic variation can be explained by climate independent of geography across both steep and shallow selection gradients. I also find that there is regional variation in the factors that drive adaptive divergence across the species range, and that it is important to consider the geographic scale of adaptation in similar studies.

References

Bachmann JC (2017) Adaptive Divergence across an Elevational Gradient in the Common Frog (Rana temporaria). University of Zurich.

Bay RA, Harrigan RJ, Le Underwood V *et al.* (2018) Genomic signals of selection predict climate-driven population declines in a migratory bird. *Science*, **359**, 83–86.

Fitzpatrick MC, Keller SR (2015) Ecological genomics meets community-level modelling of biodiversity: mapping the genomic landscape of current and future environmental adaptation. *Ecology Letters*, **18**, 1–16.

Hermisson J, Pennings PS (2005) Soft sweeps: molecular population genetics of adaptation from standing genetic variation. *Genetics*, **169**, 2335–52.

Hewitt G (1999) Post-glacial re-colonization of European biota. *Biological Journal of the Linnean Society*, **68**, 87–112.

Laugen AT, Laurila A, Rasanen K, Merilä J (2003) Latitudinal countergradient variation in the common frog (Rana temporaria) development rates – evidence for local adaptation. *Journal of Evolutionary Biology*, **16**, 996–1005.

Lenormand T (2002) Gene flow and the limits to natural. *Trends in ecology & evolution*, **17**, 183–189.

Meester L De, Vanoverbeke J, Kilsdonk LJ, Urban MC (2016) Evolving Perspectives on Monopolization and Priority Effects. *Trends in Ecology & Evolution*, **31**, 136–146.

Miaud C, Merilä J (2001) Local adaptation or environmental induction? Causes of population differentiation in alpine amphibians. *Biota*, **2**, 31–50.

Orsini L, Mergeay J, Vanoverbeke J, De Meester L (2013) The role of selection in driving landscape genomic structure of the waterflea Daphnia magna. *Molecular ecology*, **22**, 583–601.

Palo JU, O'Hara RB, Laugen AT *et al.* (2003) Latitudinal divergence of the common frog (Rana temporaria) life history traits by natural selection: evidence from a comparison of molecular quantitative genetic data. *Molecular Ecology*, **12**, 1963–1978.

Parmesan C (2006) Ecological and Evolutionary Responses to Recent Climate Change. *Annual Review of Ecology, Evolution, and Systematics*, **37**, 637–669.

Pennings PS, Hermisson J (2006) Soft sweeps II - Molecular population genetics of adaptation from recurrent mutation or migration. *Molecular Biology and Evolution*, **23**, 1076–1084.

Pritchard JK, Di Rienzo A (2010) Adaptation - not by sweeps alone. *Nature Reviews Genetics*, **11**, 665–7.

Rellstab C, Fischer MC, Zoller S *et al.* (2016) Local adaptation (mostly) remains local: reassessing environmental associations of climate-related candidate SNPs in Arabidopsis halleri. *Heredity*, 1–9.

Rellstab C, Gugerli F, Eckert AJ, Hancock AM, Holderegger R (2015) A practical guide to environmental association analysis in landscape genomics. *Molecular Ecology*, **24**, 4348–70.

Sillero N, Campos J, Bonardi A *et al.* (2014) Updated distribution and biogeography of amphibians and reptiles of Europe. *Amphibia-Reptilia*, **35**, 1–31.

Sork VL, Aitken SN, Dyer RJ *et al.* (2013) Putting the landscape into the genomics of trees: approaches for understanding local adaptation and population responses to changing climate. *Tree Genetics & Genomes*, **9**, 901–911.

Stapley J, Reger J, Feulner PGD *et al.* (2010) Adaptation genomics: the next generation. *Trends in ecology & evolution*, **25**, 705–12.

Vitti JJ, Grossman SR, Sabeti PC (2013) Detecting Natural Selection in Genomic Data. *Annu. Rev. Genet*, **47**, 97–120.

Yeaman S (2015) Local Adaptation by Alleles of Small Effect. *The American Naturalist*, **186**, S74–S89.

European common frog (*Rana temporaria)* recolonised Switzerland from multiple glacial refugia in northern Italy via trans- and circum-Alpine routes

The high mountain ranges of western Europe have had a profound effect on the recolonisation of Europe from glacial refugia. The Alps present a particularly intersting case, because they present an absolute barrier to dispersal for most lineages, obstructing recolonisation from multiple refugia in the Italian Alps. Here I investigate the effect of the European Alps on the phylogeographic history of *Rana temporaria* across its range in Switzerland. Based on partial *cytochrome b* and *COX1* sequences I find two mitochondrial lineages that occur roughly north and south of the alpine ridge bisecting Switzerland, with contact zones between them in eastern and western Switzerland. The northern haplogroup falls within the previously identified common western European haplogroup, while the southern haplogroup is unique to Switzerland. We find that the lineages diverged ~110 kya, approximately the onset of the last glacial maximum, which suggests that they are from different glacial refugia. Phylogenetic analyses suggest that the lineages originate from two refugia in northern Italy, and colonised Switzerland via trans- and curcum-alpine routes. Our results show that the European Alps is a semi-permeable barrier to dispersal for *R. temporaria*, and have contributed to the complex recolonisation history of Switzerland.

Introduction

Contemporary geographic distributions of genetic variation result from processes occurring on diverse time scales. For example, climate oscillations and associated ice-ages that occurred during the last ~700 ky had a major influence on contemporary European ecosystems (Taberlet *et al.* 1998; Hewitt 1999, 2000). The late Pleistocene (130-10 kya) saw dramatic changes in temperature, with glacial periods initiated by decreases in temperature of 10 - 14 °C in as little as 10-20 years, and lasting 70-5000 years (Dansgaard *et al.* 1993). The extent of the ice-sheets and inhospitable habitat restricted many species to three major refugia: the Balkans, Italy, and the Iberian Peninsula (Taberlet *et al.* 1998). After the retreat of the ice-sheets, Europe was recolonised by the surviving populations. Differing dispersal abilities, residual habitat discontinuities, and geographic complexities have determined these recolonisation routes. This has had a fundamental impact on the extant geographic distribution of species and their genetic diversity (Roy *et al.* 1996; Taberlet *et al.* 1998; Hewitt 2000; Willis & Whittaker 2008; Avise & Riddle 2009).

High mountain ranges such as the European Alps and Pyrenees were the last to deglaciate. For taxa adapted to warm climates, the persistence and extent of alpine ice-sheets and the east-west orientation of the mountain ranges hindered biotic recolonisation of Europe from their refugia (Taberlet *et al.* 1998). Cold-tolerant organisms exhibit a more complex history. Many expanded their range during glacial periods, some persisted in nunataks within the Alps, and trans-alpine recolonization has occurred in several species (Mátyás & Sperisen 2001; Lugon-Moulin & Hausser 2002; Parisod 2008; Yannic *et al.* 2008; Braaker & Heckel 2009). However, the effect of the Alps on the biogeography of cold-tolerant species remains understudied, particularly for vertebrates (but see Yannic *et al.* 2008, Braaker & Heckel 2009). Here we describe the phylogeographic history of a cold-tolerant amphibian, the European common frog (*Rana temporaria*) by densely sampling populations across the Alps in Switzerland. This species is the most widespread anuran in Europe, occurring from northern Italy and Spain to the

sub-arctic tundra in Fennoscandia in the north and the Ural mountains in the east. *R. temporaria* is ubiquitous in the European Alps and occurs up to an elevation of 2600 m (Sillero *et al.* 2014).

Two deeply diverged (~0.7 Mya) mitochondrial lineages occur in western and eastern/northern Europe, respectively (Palo *et al.* 2004). A contact zone has been identified that extends from northern Germany (Schmeller *et al.* 2008) to southern France, the northern lowlands of Switzerland, and north western Italy (Teacher *et al.* 2009). The precise location and structure of the contact zone in the Alps has not been resolved by existing studies (Stefani *et al.* 2012; Rodrigues *et al.* 2013), although this will be particularly interesting because it reflects the phylogeographic history and the distribution of genetic diversity in this region. The western haplogroup has higher genetic diversity than the eastern haplogroup. This suggests that *R. temporaria* recolonised western Europe from multiple glacial refugia. Both Iberia (Teacher *et al.* 2009) and Italy (Stefani *et al.* 2012) have been suggested as the main refugia for the western haplogroup. However, sampling has not been available to reconstruct the main routes of post-glacial recolonisation of the western haplogroups, and it remains unclear whether the Alps was a barrier to re-colonisation into western Europe.

Our main aims are to determine whether 1) a contact zone exists between Eastern and Western haplotypes in Switzerland, 2) Switzerland was colonised from one or multiple glacial refugia, and 3) determine whether the Alps was a barrier to colonisation.

The geography of the Alps between northern Italy and Switzerland are characterised by two main high mountain ridges running roughly east to west. We refer to the first as the southern Alpine ridge. This comprises the highest peaks and roughly constitutes the border between Italy and Switzerland. The second ridge is called the northern Alpine ridge, and roughly bisects Switzerland into a northern and southern part.

Materials and Methods

Sampling

To establish phylogeographic patterns, we sampled eggs from 2-9 freshly-laid clutches in 72 populations of *R. temporaria* in the Swiss Alps during the 2013 breeding season. Locations, elevations, and sample sizes of the populations are given in Table S1. The eggs were transported to the University of Zürich, where they were reared at 23-25 °C until they hatched and reached stage 36 (Gosner, 1960). One individual from each clutch (*N* = 368) was euthanised and stored in ethanol.

Molecular lab work

Total DNA was extracted from tadpole tails by overnight digestion in 10% proteinase K solution at 56 °C and extracted using the Qiagen Biosprint 96 DNA blood kit (Qiagen, CA, USA). DNA was eluted in 100 - 200 µL buffer AE (QIAgen).

To establish phylogeographic patterns at a European scale, we sequenced the same section of the *cytb* gene that was studied in earlier work on *R. temporaria* (Teacher *et al.* 2009; Vences *et al.* 2013). A 448bp fragment of *cytb* was amplified by PCR using *Rana-cytb-F2* (5' TTAGTAATAGCCACAGCTTTTGTAGGC), and *Rana-Cytb-R2* (5' AGGGAACGAAGTTTGGAGGTGTGG) primers (Vences *et al.* 2013) with an annealing temperature of 53 °C.

The aforementioned studies do not include fine-scale sampling of the Italian Alps. To determine the phylogenetic relationship between Swiss and Italian samples generated sequence data to compare with Italian samples from Stefani *et al.* (2012). For a subset of 44 populations (151 families), we PCR-amplified a 628bp fragment of COXI. Primers L4437 (5' AAGCTTTCGGGCCCATACC) and H6564 (5' GGGTCTCCTCCTCCAGCTGGGTC) with an annealing temperature of 49 °C.

Amplified COXI and *cytb* fragments were subjected to Sanger sequencing after we purified them using a standard ExoSap protocol: 0.25 µl Exonuclease I (20 U/µl; New England Biosystems), 0.5 µl rAPID Alkaline

Phosphatase (SAP) (1 U/μl), 7.25 μl nuclease free water, 8 μl PCR product. Cycling conditions were as follows: 37 °C for 45 min, 80 °C for 15 min, and held at 10 °C. From the cleaned product, 2 μl per sample was used in 10 μl sequencing reactions (BigDye Sequencing Kit, Applied Biosystems). Samples were sequenced in one direction for each gene using an automated 3130xl DNA Analyzer (Applied Biosystems, CA, USA) and the sequences were aligned using CLC Main Workbench 5.0.2 (CLC Bio) and BioEdit (Hall 1999). Sequences were verified as *cytb* and COXI using the BLASTN 2.2.24 algorithm (Altschup *et al.* 1990) implemented on the National Center for Biotechnology Information (NCBI) online platform.

Palo *et al*. (2004) identified a restriction enzyme cut site at position 277 in the *cytb* sequence that distinguishes between the eastern and the western mitochondrial clades. We used this assay to assign our samples to clades. Restriction digests were performed using 8 μl of the PCR product, 0.1 μl 10000 U/ml *StyI* restriction enzyme (New England Biolabs), 0.9 μl water, and 1 μl 10X Buffer (Qiagen). PCR products were digested for 30 minutes at 37 °C, where after they were resolved on a 1% agarose gel and assessed by eye under UV-light.

Analysis

Geographic distribution of genetic variation. − To visualise the geographic distribution of genetic variation, haplotype networks were constructed with *cytb* and COX1 sequences using statistical parsimony (Templeton *et al.* 1992) as implemented in TCS version 1.21 (Clement *et al.* 2000). Within Switzerland, data generated in this study were used: 368 individuals sequenced at 448bp *cytb*, and 151 individuals sequenced at 1076bp (628bp COX1 and 448bp *cytb*). At a broader geographic scale, we visualized genetic variation by combining our data with those from elsewhere in Europe obtained from NCBI (accession numbers for *cytb*: KC799522.1 - KC800122.2; Vences *et al.* 2013). This dataset comprised 969 individuals sequenced at 331bp cytb. Note that these data do not include Italian samples described in Stefani *et al.* (2012), as a different *cytb* fragment was amplified in their study. A separate analysis was conducted based on COXI to determine

the phylogenetic relationship between Swiss and Italian samples (see below).

Phylogenetic analyses. − Evolutionary relationships among geographic groups in the haplotype network were determined with phylogenetic analyses on three datasets: 1) *cytb1*, 2) COXI, 3) combined *cytb1* and COXI. All datasets comprised of data from this study and all available data from Vences *et al.* (2013). Outgroup species were chosen based on the *Rana* phylogeny in Veith *et al.* (2003) and data available from NCBI. For *cytb*, the outgroups were *R. pyrenaica* (KC799521.1), *R. arvalis* (KC800123), *R. iberica* (KC799485), and *R. italica* (KC799493). For COXI, the outgroups were *R. pyrenaica* (KC977251.1) and *R. arvalis* (JN971596.1). As a first step we identified the evolutionary model that best fit each dataset, and secondly we determine the most likely phylogenetic tree under the chosen model. Phylogenetic trees were estimated using both Bayesian and likelihood methods.

The most likely evolutionary mutation model was determined for each dataset by comparing 88 alternative models in jModeltest2 (Darriba *et al.* 2012) and using the Akaike Information Criterion (Akaike 1973) to select the model best supported by the data.

Phylogenetic analyses were conducted using maximum likelihood and Bayesian methods. Maximum likelihood analyses were implemented in RAxML v.8.2.2 (Stamatakis 2006). For each dataset, 1000 rapid bootstrap replicates were conducted followed by a maximum likelihood search for the most likely phylogenetic tree. Bayesian phylogenetic analyses were conducted using MrBayes v.3.2.1 (Huelsenbeck & Ronquist 2001). Analyses were run for 1×10^6 generations and sampled every 1000 generations. Four Metropolis-coupled Markov chain Monte Carlo (MCMCMC) chains were used for two replicate runs of each dataset, with chain heating kept at default temperatures. Convergence was assessed using the diagnostic output from the *sump* command: Adequate effective samples sizes (ESS) for each parameter (>200), appropriate mixing of chains, and an average standard deviation of split frequencies between independent runs <0.05. A burnin of 25% was used to obtain the consensus phylogram and posterior probabilities for each bipartition.

Switzerland is bisected by an east-west alpine ridge, which appears as a

dividing line for the two major lineages we detected in each haplotype network. We will refer to these lineages as CH-North (northern Swiss lineage) and CH-South (southern Swiss lineage). A third lineage referred to as CH-brown is phylogenetically closer to the CH-North lineage, but occurs exclusively south of bisecting alpine ridge. To estimate the geographic distribution of genetic diversity, haplotype diversity (H) and nucleotide diversity (π) were calculated in DnaSP 5.0 (Librado & Rozas 2009) for all Swiss populations and for CH-North and CH-South separately.

We estimated the coalescence date for all *R. temporaria* haplotypes and the Swiss subset of haplotypes using a Bayesian Markov chain Monte Carlo analysis with *R. arvalis* and *R. pyrenaica* as outgroups (BEAST v. 1.8.2; Drummond *et al.* 2012). COX1 and *cytb1* were analysed as a single dataset, totaling 1076bp, because jModeltest inferred a similar nucleotide substitution model for both gene fragments. We used the GTR + I substitution model, with estimated base frequencies, and 6 rate categories. Priors were set to their defaults except for the date of divergence of the three *Rana* species, which was specified as a lognormal distribution with a mean of 3.18 Mya and standard deviation of 0.43 Mya (Veith *et al.* 2003). Tests were run with all demographic and molecular clock combinations as priors, and the most likely model was selected based on a marginal likelihood test. The final analysis was run using a relaxed molecular clock, and a demographic model of constant population size.

To test the hypothesis that Switzerland was colonised from the Italian refugium, we estimated the phylogenetic relationship between 569 bp COX1 data from this study, and data available on NCBI. This analysis was based on COX1 data only, because the publicly available *cytb* fragment sequenced for samples from northern Italy (Stefani *et al.* 2012) does not overlap with that of samples from Europe (Vences *et al.* 2013). We included *R. temporaria* haplotypes from the Italian Alps (Stefani *et al.* 2012; FN813783-FN813812), Europe (Vences *et al.* 2013; KC977228.1 - KC977251.1) and outgroups *R. pyrenaica* (KC977251.1) and *R. arvalis* (JN971596.1). Phylogenetic analyses were conducted as described above.

Demographic history. − To test for signals recent changes in population size we used Tajima's D, Fu's F, and the DNA sequence mismatch distribution (Rogers & Harpending 1992). A mismatch distribution tests for signatures of population expansion based on the number of nucleotide differences (mismatches) between DNA sequences. When population size remains stable the distribution of mismatches is expected to be multimodal, reflecting the stochastic nature of gene trees. When populations expand, we expect a unimodal distribution of mismatches. Signals of population expansion were tested in the haplogroups using Harpending's raggedness index (RI). The expansion time was estimated under a model of pure demographic expansion with parameters set to default in Arlequin (Excoffier & Lischer 2010). The parameter of demographic expansion, Tau, was estimated according to Schneider & Excoffier (1999). The validity of the expansion model was tested using the sum of square deviations (SSDs) between the observed and expected mismatches as implemented in Arlequin.

Results

The final dataset comprised 368 families from 72 populations sequenced at 448bp of the *cytb* gene, and 151 families from 44 populations sequenced at 628bp of COXI.

Identification of eastern and western clades

Based on the absence of the diagnostic *cytb* restriction site, all Swiss samples were determined to be from the western clade.

Geographic distribution of genetic variation

Haplotype network - Overall the haplotype network shows an eastern (blue) and western (other colours) clade with much higher genetic diversity found in the western clade (Fig. 2.1). The Swiss samples assigned to three western haplogroups; a green haplogroup which is widespread across

western Europe and Great Britain, a brown haplogroup found almost exclusively in the Pyrenees (Vences et al. 2013), and a Swiss specific haplogroup shown in purple described here for the first time. Within Switzerland the green haplogroup is found north of the Alpine ridge bisecting Switzerland, while the purple and brown haplogroups occur south of the ridge (Fig. 2.2). This division was confirmed when Swiss samples were analysed using cytb alone and in combination with COXI (Fig. S2.1). The haplogroup found north of the Alps (CH-North) corresponded with the widespread western European haplogroup, which is thought to have colonised western Europe from the Pyrenees (green haplotypes in Fig. 2.1; Vences *et al.* 2013). The brown haplogroup (CH-brown) was closely related to the common western European haplogroup, and occurs almost exclusively in the Pyrenees (brown haplotypes in Fig. 2.1). The southern haplogroup was unique to Switzerland, and fell between the haplogroups sampled in the Pyrenees and those sampled in Croatia and Greece in the haplotype network. CH-North and CH-South co-occurred in some populations but across all elevations in eastern Switzerland and at low elevation in western Switzerland. A second western European lineage (CH-brown: brown haplotypes in Fig. 2.1), found in the Pyrenees and restricted roughly to the southern part of Europe, co-occurred with CH-South haplotypes in southern Switzerland (Fig. 2.2).

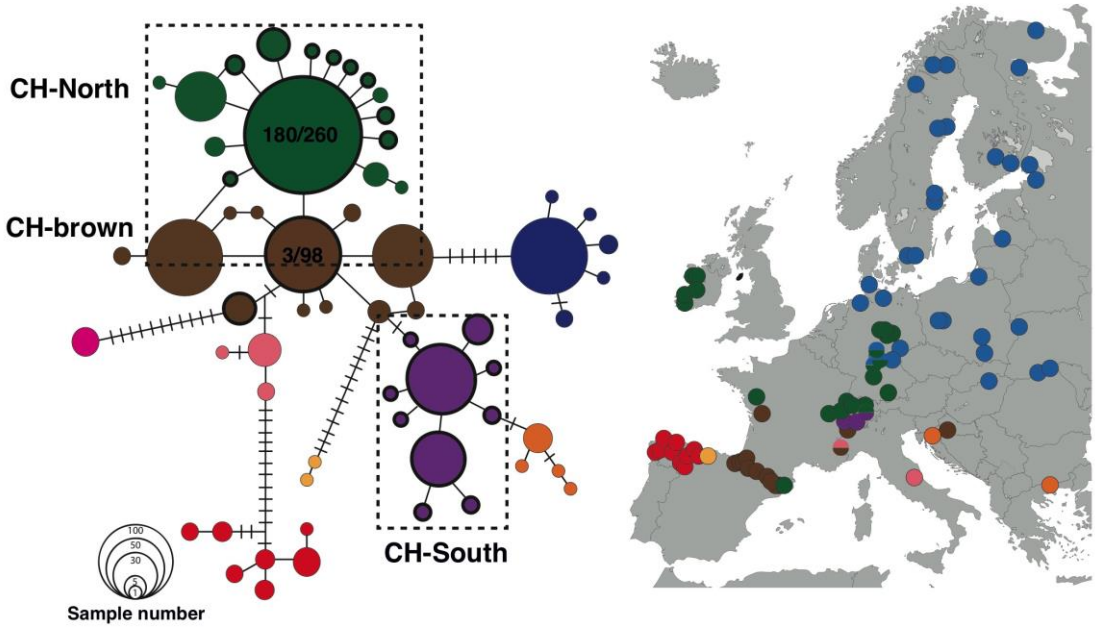


Figure 2.1 Haplotype network of 331bp of the *cytb* gene sequenced for this study and combined with data from Vences *et al.* (2013). Geographic distribution of the haplotypes within the eastern (blue) and western (other colours) clades is shown on a map of Europe. Swiss haplotypes (shown with bold outlines) fall within the green (CH-North) and brown (CH-brown) haplogroups, and a Swiss-specific purple haplogroup (CH-South). Numbers in the green and brown haplotypes show the proportion of samples from this study assigning to that haplotype.

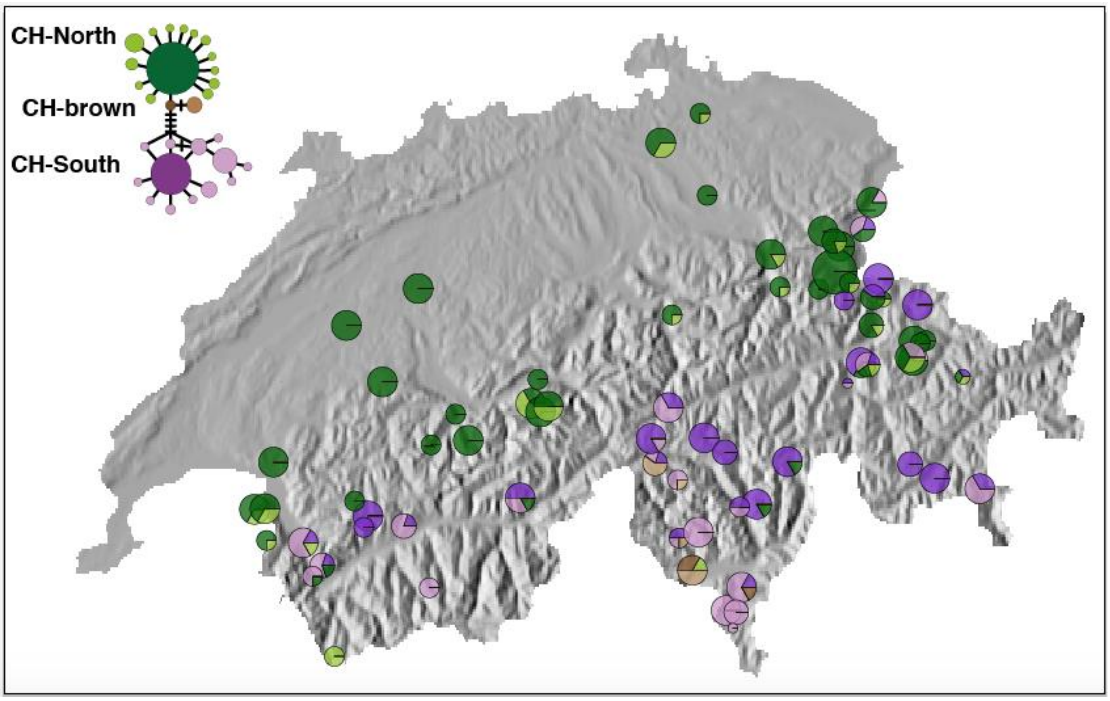


Figure 2.2 Swiss *cytb* (448bp) haplotypes are distributed roughly north (CH-North) and south (CH-South and CH-brown) of the northern Alps. Pie charts show the proportion of individuals that had northern or southern haplotypes in each population, with pie chart size reflecting the number of samples per population. The inset depicts the *cytb* haplotype network calculated for all Swiss samples (368 individuals from 72 populations sequenced at 448 bp). The most common haplotype within each haplogroup is shaded in a darker colour compared to the derived haplotypes. Thus the distribution of the common and derived haplotypes are shown with pie charts on the map.

Phylogenetic analyses - The same phylogenetic relationships were recovered using ML and Bayesian approaches, as well as between the *cytb1*, COXI, and concatenated datasets, and were concordant with previously published phylogenies (Fig. 2.3) (Veith *et al.* 2002; Vences *et al.* 2013). A basal split in the *R. temporaria* clade separates three Spanish clades that occur exclusively in Iberia (Vences *et al.* 2013) from the rest of the samples. *R. temporaria* is further split into an eastern and a western clade with high posterior probability (0.99). Within the western clade, a further split is found between populations from southern Switzerland and all other western populations (including northern Switzerland/western Europe; posterior probability = 0.97).

The estimated mean coalescent times in Fig. 2.3 support a divergence between the eastern and western clades at ~210 kya, and a more recent

divergence between CH-South and the rest of the western clade at ~110 kya. The most likely coalescent model based on marginal likelihood tests was a relaxed molecular clock and constant population size. The effective sample size (ESS) for all parameters was above 200, and all independent runs showed convergence and stationarity.

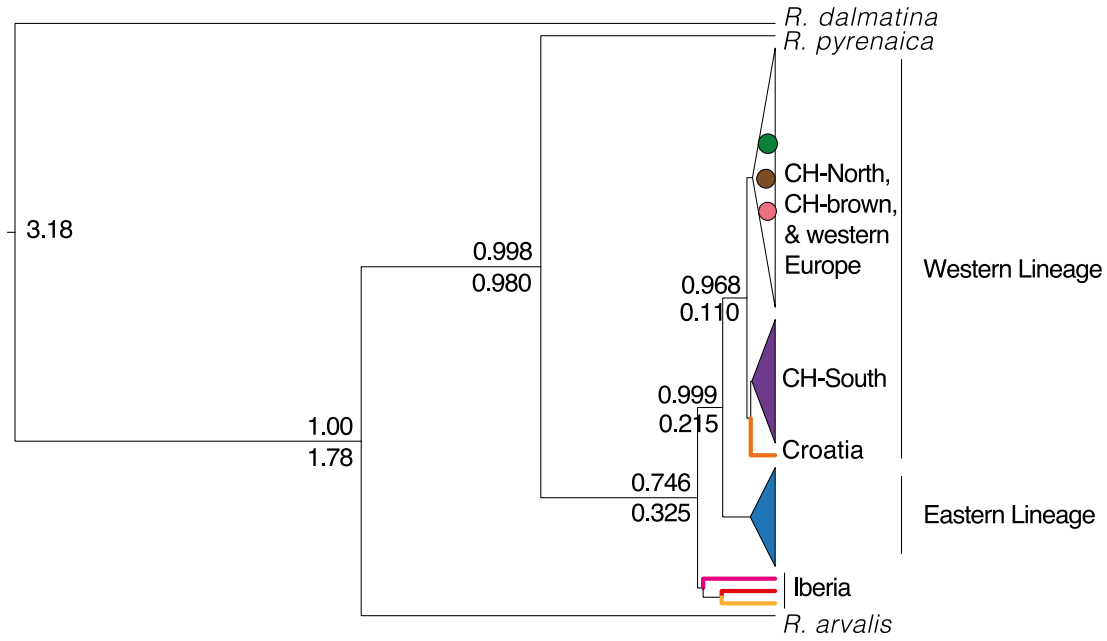


Figure 2.3 Bayesian phylogenetic tree of concatenated *cytb1* and COXI genes**.** The eastern and western lineage, the CH-South, CH-North, and CH-brown haplogroups are defined as in Fig. 2.1. The eastern and CH-South haplogroups are coloured as in Fig. 2.1. CH-North, CH-brown and western European haplotypes are indicated by green, brown, and pink circles (colours as in Fig. 2.1). Posterior probability is shown above the line, and the inferred coalescent time (Mya) as determined by BEAST is shown below the line. The calibration used (3.18Mya; Vences *et al.* 2003) is shown at the root of the tree.

Distribution of genetic variation - Genetic diversity within Switzerland was higher in populations found south of the alpine ridge bisecting Switzerland. Haplotype and nucleotide diversity based on 448bp *cytb* gene fragment was roughly double in CH-South when compared to CH-North populations, and roughly three times higher based on 628bp COX1 sequence (Table 2.1; diversity measures for individual populations are reported in Tables S2.1 and S2.2).

Demographic history

The mismatch distribution analyses did not converge for CH-North for any of the datasets, but Tajima's D and Fu's F were negative and significant for all CH-North datasets indicative of a population expansion. Tajima's D and Fu's F were non-significant for CH-South, and there was no evidence for population expansion for any of the CH-South analyses.

Origin of haplotypes in Switzerland

Phylogenetic analysis based on 569 bp of COX1 revealed that haplotypes from the Italian Alpine Lineage (AP) I described in Stefani *et al.* (2012) grouped with CH-South haplotypes with high posterior probability (0.94; Fig. 2.4). Three haplotypes identified by Stefani *et al.* (2012) − FN813810.1, FN813785.1, FN813786 − were identical to CH-South haplotypes. API occurs in the central Italian Alps and contains the most abundant Italian Alpine haplotype (CA2), which was identical to CHS03. In our sample, this haplotype was found only in a population on a high pass in southern Switzerland, on the border with Italy. The most widespread CH-South haplotype was identical to the widespread API haplotype DE10. A third haplotype (API: VF6) was identical to CHS06, found only at high elevation in southwestern Switzerland (Figs. 2.4 & 2.5).

The northern Swiss lineage (CH-North) clustered with APIII with high posterior probability (1.0). One previously identified haplotype from APIII (FN813809.1 = VF3) was identical to the most common CH-North haplotype (CHN01). Geographically, APIII and VF3 are found in northwestern Italy, but in Switzerland CHN01 occurs mostly north of the bisecting alpine ridge (Fig. 2.5). APIII clusters with a haplotype from southern France with high posterior probability (0.93; Fig. 2.4), which suggests that this haplogroup is analgous to the green haplogroup described by Vences *et al.* (2013) and is widespread across France, Germany, and Great Britain.

Table 2.1 Diversity measures and neutrality tests for the *cytb*, COX1, and concatenated datasets generated in this study. CH-North includes all samples with haplotypes from the Northern haplogroup, and CH-South from the Swiss-specific haplogroup. N = Number of samples; length = sequence length in base pairs; nr Hap = the number of haplotypes detected, Hd = haplotype diversity (standard deviation); Nd = nucleotide diversity (standard deviation). Three tests for selection are included. * = $P < 0.05$; ** = $P < 0.01$; *** = $P < 0.001$.

Pop Group	N	length (bp)	nr Hap	Hd (SD)	π (SD)	Tajima's D	Fu & Li's D	Fu & Li's F
				cytb				
All	368	448	29	0.712 (0.020)	0.00857 (0.00019)	-0.377	-2.896*	-2.202
CH-North	217	448	17	0.347 (0.042)	0.00126 (0.00021)	-2.207**	-3.945**	-3.918**
CH-South	151	448	12	0.633 (0.034)	0.00249 (0.00016)	-1.068	-1.429	-1.550
				COX1				
All	151	628	15	0.614 (0.033)	0.00517 (0.00023)	-0.1242	-1.797	-1.381
CH-North	92	628	7	0.167 (0.053)	0.00031 (0.00011)	-2.039*	-3.117*	-3.264**
CH-South	59	628	8	0.485 (0.078)	0.00103 (0.00021)	-1.468	-1.179	-1.497
				concatenated cytb + COX1				
All	141	1076	31	0.792 (0.029)	0.00658 (0.00026)	-0.067	-2.848	-2.034
CH-North	86	1076	17	0.557 (0.065)	0.00080 (0.00014)	-2.209***	-3.415**	-3.545**
CH-South	55	1076	14	0.706 (0.064)	0.00157 (0.00018)	-1.194	-2.147	-2.158

***** p <0.05, ****** p <0.02, ******* p<0.0

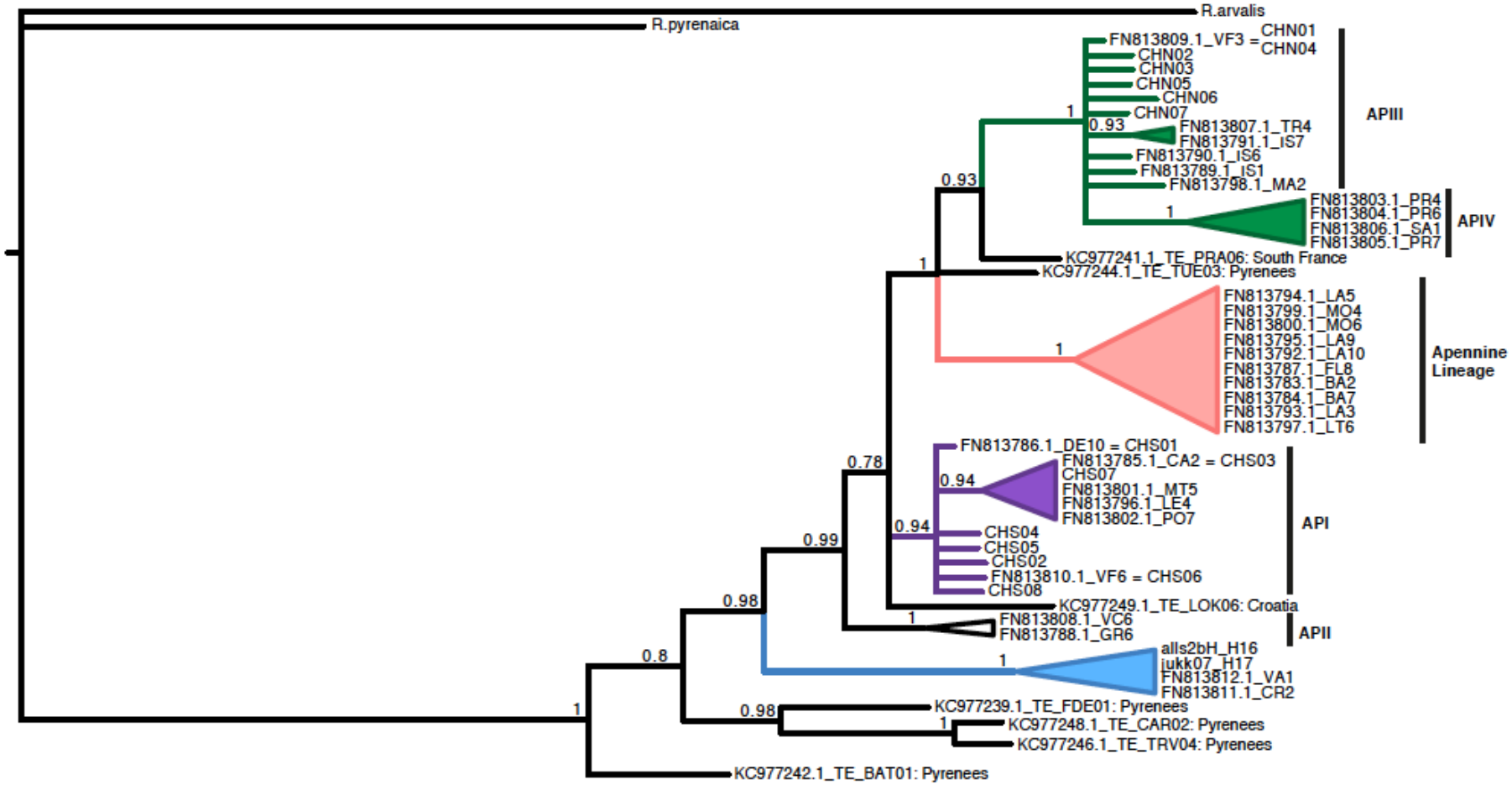


Figure 2.4 Bayesian phylogenetic tree of COX1 haplotypes (569 bp) sampled across the *R. temporaria* range. Italian lineages described by Stefani *et al.* (2012) (Alpine lineage I-IV and Apennine lineage; NCBI code FNxx) are shown with vertical black bars. Four haplotypes from this study match Stefani *et al.* (2012) haplotypes and are indicated with an equal (=) sign. Coloured regions match the haplogroups identified in Fig. 2.1. Haplotypes from this study are indicated with CHN (CH-North) and CHS (CH-South). CH-brown haplotypes are not visible in this figure as these populations were not sequenced at COX1. Haplotypes from Vences *et al.* (2013) are shown with their NCBI codes (KCxx) and their sample locations. Posterior probability of >0.5 is shown above each node

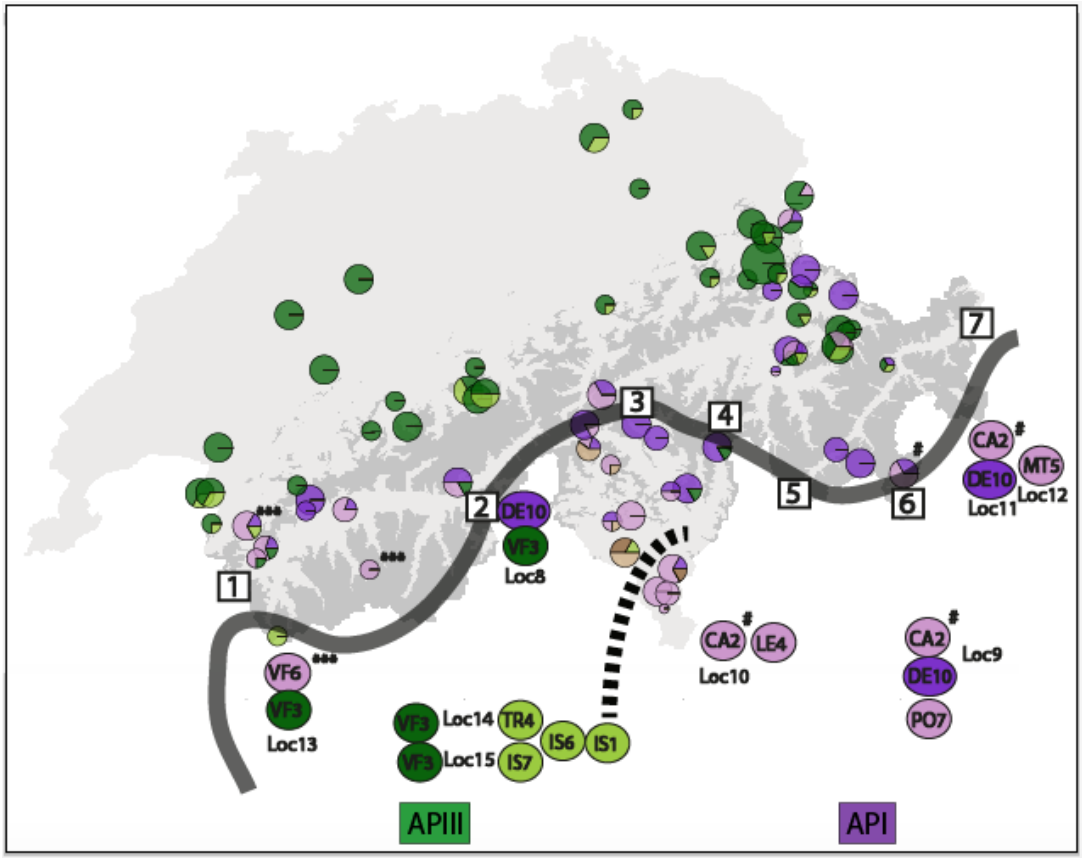


Figure 2.5 Possible trans-alpine colonisation routes from the Italian Alps into Switzerland. The map shows elevation >2000 m in dark grey; thus the light grey highlights the potential regions available for colonising populations. The thick solid grey line roughly denotes the central Alpine ridge. Haplotype colours correspond with the haplogroups identified in Fig. 2.2 and 4. We show the geographic distribution of the COX1 haplotypes identified by Stefani *et al.* (2012) in Northern Italy bordering Switzerland. The dashed line indicates the Ticino valley that roughly separates API (light and dark purple) from APIII (light and dark green). For illustration purposes we the haplotypes shown in Switzerland are from *cytb* sequences (i.e. Fig. 2.3) since mitochondrial haplotypes are linked and this gene was sequenced more geographically extensively than COX1 for our study. The Italian populations are defined as in Stefani *et al* (2012); Loc8: Pramollo (46.36 N 11.86 E), Loc9: M.Pora (45.88 N 10.1 E), Loc10: Lemna (45.85 N 9.18 E), Loc11: S.Caterina (46.41 N 10.49 E), Loc12: V.Martello (46.49 N 10.69 E), Loc13: V.Ferret (45.85 N 7.02 E), Loc14: Traversella (45.53 N 7.72 E), Loc15: Issiglio (45.46 N 7.32 E). Four haplotypes matched sequences from this study and their geographic locations are shown as follows: VF3: Dark green; DE10: Dark purple; CA2:#; VF6:***. The numbered blocks are mountain passes less than 2000 m elevation presenting possible cross-alpine colonisation points: 1=Col de Forclaz, 2=Simplonpass, 3=Lukmanierpass, 4=San Bernardino Pass, 5=Malojapass, 6=Bernina Pass , 7=Reschenpass.

Discussion

We interpret our results to indicate that *Rana temporaria* recolonised western Europe almost exclusively from multiple refugia in the Italian Alps. Populations originating from hypothesised refugia in the Iberian Peninsula occur only in Spain and southwestern France. Our fine-scale sampling of populations across Switzerland provides evidence of a trans-alpine colonisation from the central Italian Alps into Switzerland, and recolonisation from the western Italian Alps around the alpine arc into much of western Europe. We found two deeply diverged mitochondrial lineages occurring roughly north and south of the northern Alps, with contact zones in eastern and western Switzerland. These results strongly suggest that that the Alps did not present a significant barrier to dispersal for this species during the recolonisation of Europe. This has resulted in complex regional variation in the genetic structure of *R. temporaria* across Switzerland, which could have important implications for population persistence and adaptive potential.

Eastern and Western European lineages

Consistent with previous research, we found a deep phylogenetic split between eastern and western European lineages of *R. temporaria* (Palo *et al.* 2004). The age of the east-west divergence, calibrated based on the *R. dalmatina* divergence time (3.18 Mya; Veith *et al.* 2003), was placed in the late Pleistocene (~0.215 Mya). The divergence date was previously estimated as ~0.71 Mya (CI 0.53-0.95 Mya) based on *cytb* sequences (Palo *et al.* 2004), corresponding to the onset of strong Pleistocene climatic cycles. The discrepancy may be attributed to the large confidence intervals associated with the statistical priors in our analyses, as well as the slightly different calibration points included in both studies. However, our estimated date and that of Palo et al.(2004) both suggest that the divergence between the eastern and western clades is attributable to the dramatic climatic cycles that occurred during the Pleistocene, which began ~1.8Mya.

Two previous studies disagreed about the existence of a contact zone between the eastern and western lineages in northern Switzerland and along the French Mediterranean into Spain (Teacher *et al.* 2009, Rodrigues et al.

2013). Our thorough sample of northern Switzerland, including several populations in the lowlands where the contact zone was said to lie (Teacher et al. 2009), uncovered no eastern haplotypes whatsoever and hence no evidence of contact between clades. Presumably, a secondary contact zone between the eastern and western haplogroups occurs somewhere farther to the east.

Post-glacial recolonisation of western Europe

Several studies have identified multiple glacial refugia in the Italian peninsula, that consequently lead to multiple deeply diverged lineages (Canestrelli *et al.* 2006, 2008; Crottini *et al.* 2007; Canestrelli & Nascetti 2008). Our results support the assertion that *R. temporaria* re-colonised western Europe from multiple sub-refugia within the Italian refugium (Stefani *et al.* 2012). We identified two western mitochondrial lineages distributed roughly north and south of the northern Alps within Switzerland. The two lineages are estimated to have diverged ~110 kya, at the onset of the most recent glacial cold period. Four Alpine lineages have been described in northern Italy, two of which (API and APIII) are distributed south of Switzerland roughly east and west of the Ticino valley (Fig. 2.5). Our phylogenetic analyses suggest that the southern (CH-South) haplogroup was most likely colonised by API from the central Italian Alps, while the northern (CH-North) haplogroup was colonised by APIII.

A thorough geographic analysis of the CH-North and CH-South lineages suggests that they recolonised western Europe via two different routes (Fig. 2.6). The geographic distribution of the southern lineage across the central Alpine ridge is evidence of multiple trans-alpine colonisations into Switzerland (Fig. 2.5). Three CH-South haplotypes were identical to AP1 haplotypes, and their geographic distribution can be used to narrow down colonisation routes. Assuming a conservative maximum habitable elevation of 2000 m (as in Braaker & Heckel 2009), there are seven possible passes across the central Alpine ridge (Fig. 2.5). The most widespread AP1 haplotype (CA2) was identical to a CH-South haplotype found in south-eastern Switzerland, which suggests colonisation across the Maloja Pass (5,

Fig. 2.5) or Bernina Pass (6, Fig. 2.5).

Southwestern Switzerland contained several sites with secondary contact between API and APIII, which could suggest a southern trans-alpine colonisation route of APIII into Switzerland. Possible routes include the Col de Ferdaz (1, Fig. 2.5) in the west or Simplonpass (2, Fig. 2.5), but the most likely point of colonization is difficult to identify because the most common CH-South haplotype (API haplotype DE10) was geographically widespread in northern Italy.

In other European taxa, studies of trans-alpine colonisation have identified the same trans-Alpine colonisation routes for both plants and animals (Lugon-Moulin & Hausser 2002; Yannic *et al.* 2008, 2012; Braaker & Heckel 2009; Parisod 2008). The example of European white oaks (Mátyás & Sperisen 2001), which presently occur only rarely above 1200 m in the Alps (Rigling *et al.* 2013), illustrates that trans-Alpine recolonization was not limited only to taxa adapted to cold climates. Presumably, dispersal over Alpine passes was more extensive when the treeline was ~200 m higher than it is today, which occurred regularly during the period of 5'000-10'000 years ago (e.g. Ammann *et al.* 2000; Tinner & Theurillat 2003; Heiri *et al.* 2006). Exact colonisation routes could be investigated with the inclusion of nuclear markers and fine-scale geographic sampling of populations across the putative passes used for colonisation.

CH-North haplotypes appear to have colonized Switzerland from the west probably up the Rhone River or from the north west over the Jura Mountains (Fig. 2.6). The geographic distribution of the derived CH-North haplotypes towards southern Switzerland (light green haplotypes; Fig. 2.2) and overall low genetic diversity (Table 2.1) suggest that this lineage expanded southwards in Switzerland rather than colonising via trans-alpine routes as with the CH-South lineage. Three possible origins of this haplogroup

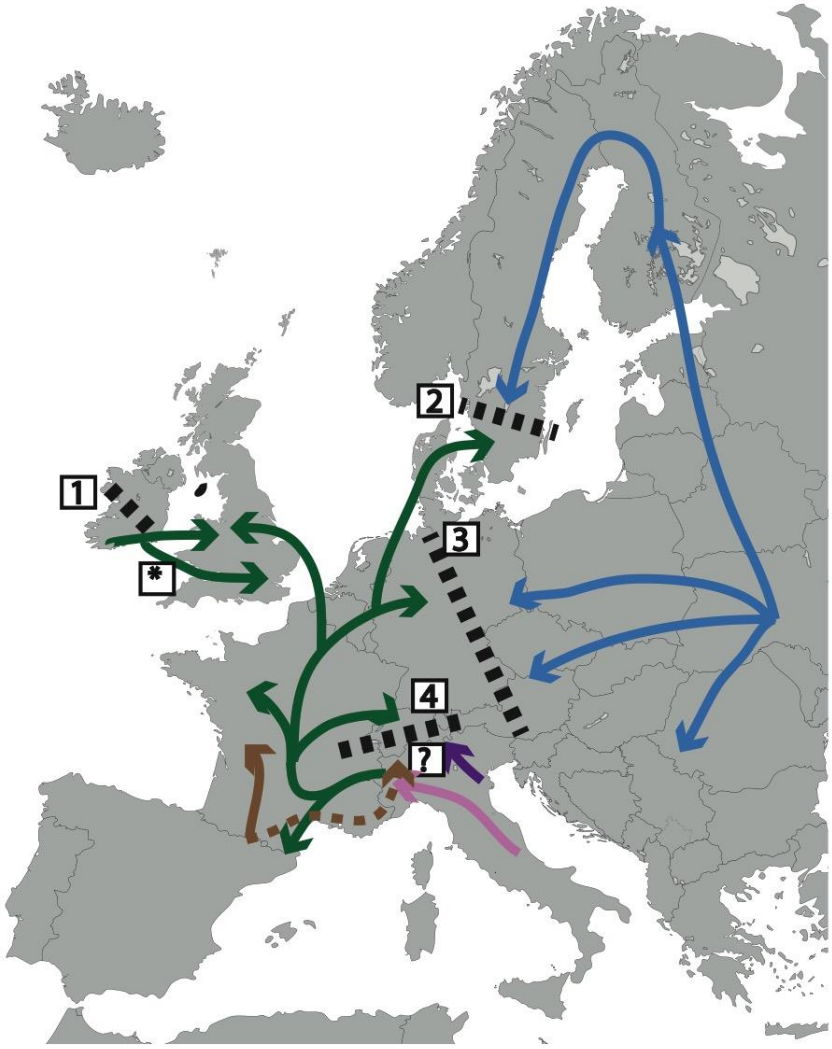


Figure 2.6 Proposed *R. temporaria* post-glacial recolonisation routes. Coloured arrows indicate the proposed origin and colonisation routes of five haplogroups. Our results suggest that Italy is the major glacial refugia for *R. temporaria*, with the common green haplogroup colonising western Europe from the north western Italian Alps, and a transalpine colonisation of the purple haplogroup into Switzerland. Dashed lines show probable regions of secondary contact between diverged lineages. 1=common western European lineage meets *R. temporaria* from Irish refugium (Teacher *et al.* 2009); 2=Likely original contact zone between eastern and western haplogroup, but only the western haplogroup is currently found here (Palo *et al.* 2004)**;** 3=Conact zone between eastern and western haplogroup in northern Germany (Schmeller *et al.* 2008), but contact zone south of this site still needs to be described; 4=contact zone between CH-South and CH-North described in this paper; ?=Possible contact zone between the Iberian (CH-brown) and Italian lineages described in this paper. The asterist (*) denotes an alternative origin and colonisation route of the green haplotype from the Irish refugium into western Europe.

have been proposed: Iberia, a cyptic refugium in southern Ireland, and northern Italy (Teacher *et al.* 2009; Stefani *et al.* 2012; Vences *et al.* 2013). Although we cannot definitively rule out Iberia or Ireland, our results seem to suggest an Italian origin. Specifically, the previously identified common western European haplogroup (T4; Vences *et al.* 2013) occupies northern Switzerland, and we found that this haplogroup is identical to Italian APIII found in the north western Italian Alps. In addition, the rest of the CH-North haplotypes form a monophyletic group with APIII (Fig. 2.4). Finally, genetic diversity is expected to decrease with distance from glacial refugia as populations experience bottleneck events with colonisation (Hewitt 1996). The genetic variation in the green/APII/T4 haplogroup is substantially higher in northern Italy than in Ireland or Iberia, suggesting that this was the main refugium of the haplogroup.

The identification of a second CH-brown lineage exclusively south of the central Alpine ridge in southern Switzerland suggests that there may be eastward expansion from Iberia south of the Alps (Fig. 2.2 & 2.5). This lineage is ubiquitous in the Pyrenees, and is found in southern France, northern Italy, and Croatia (Vences *et al.* 2013). Deglaciation of the Tinee Valley on the border between France and Italy occurred during the Oldest Dryas (~15kya), earlier than the rest of the Alps where glaciers retreated during the Younger Dryas (~11kya) and the late ice retreat (~9kya) (Darnault *et al.* 2011). This potential early colonisation corridor could have provided the way for an eastward expansion from the Pyrenees, across northern Italy, to Croatia. Although westward expansion from the Carpathian refugium through northern Italy has been described (Demesure *et al.* 1996; Taberlet *et al.* 1998), we found no examples of an eastward expansion from Iberia into northern Italy. Longer mtDNA fragments and more extensive sampling would be needed to investigate this unusual colonisation route further.

Conclusions

The *R. temporaria* western mitochondrial clade is exceptional in that it harbours more genetic diversity than any other European brown frogs (Palo *et al.* 2004; Vences *et al.* 2013). We show that much of this variation can be

attributed to the European Alps. The topographic complexity of the Alps has resulted in multiple glacial refugia and subsequent divergence between *R. temporaria* lineages. After the retreat of the glaciers, the Alps presented a semi-permeable barrier to dispersal for these lineages, resulting in resulting in multiple colonisation events in the south. However, this also limited colonisation from other refugia in northern Italy, which subsequently followed a circum-Alpine route to colonise western Europe and Switzerland (Fig. 2.6). This has resulted in two deeply diverged lineages that occur in the north and the south of Switzerland with contact zones in the east and west. Our work highlights how fine-scale phylogeographic studies across the Alps can elucidate of the phylogeographic history of cold-adapted species in Europe.

References

Akaike H (1973) Information theory and an extension of the maximum likelihood principle. In: *2nd International Symposium on Information Theory, Tsahkadsor, Armenia, USSR*, pp. 267–281. Budapest:Akadémiai Kiadó.

Alberto F, Niort J, Derory J *et al.* (2010) Population differentiation of sessile oak at the altitudinal front of migration in the French Pyrenees. *Molecular Ecology*, **19**, 2626–2639.

Alton LA, Franklin CE (2017) Drivers of amphibian declines: effects of ultraviolet radiation and interactions with other environmental factors. *Climate Change Responses*, **4**, DOI 10.1186/s40665-017-0034-7.

Altschup SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic Local Alignment Search Tool. *Journal of Molecular Biology*, **205**, 403–410.

Alvarez N, Thiel-Egenter C, Tribsch A *et al.* (2009) History or ecology? Substrate type as a major driver of spatial genetic structure in Alpine plants. *Ecology Letters*, **12**, 632–640.

Ammann B, Birks HJB, Brooks SJ *et al.* (2000) Quantification of biotic responses to rapid climatic changes around the Younger Dryas — a synthesis. *Paleogeography, Paleoclimatology, Paleoecology*, **159**, 313–347.

Avise JC, Riddle B (2009) Phylogeography : Retrospect and Prospect. *Journal of Biogeography*, **36**, 3–15.

Bachmann JC (2017) Adaptive Divergence across an Elevational Gradient in the Common Frog (Rana temporaria). University of Zurich.

Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, **30**, 2114–2120.

Bonin A (2008) Population genomics: a new generation of genome scans to bridge the gap with functional genomics. *Molecular ecology*, **17**, 3583–4.

Bonin A, Taberlet P, Miaud C, Pompanon F (2006) Explorative genome scan to detect candidate loci for adaptation along a gradient of altitude in the common frog (Rana temporaria). *Molecular biology and evolution*, **23**, 773–83.

Borcard D, Legendre P (2002) All-scale spatial analysis of ecological data by means of principal coordinates of neighbour matrices. *Ecological Modelling*, **153**, 51–68.

Braaker S, Heckel G (2009) Transalpine colonisation and partial phylogeographic erosion by dispersal in the common vole (Microtus arvalis). *Molecular Ecology*, **18**, 2528–2531.

Brady LD, Griffiths RA (2000) Developmental responses to pond desiccation in tadpoles of the British anuran amphibians (Bufo bufo, B. calamita and Rana temporaria). *Journal of Zoology, London*, **252**, 61–69.

Van Buskirk J (2012) Permeability of the landscape matrix between amphibian breeding sites. *Ecology and evolution*, **2**, 3160–7.

Canestrelli D, Cimmaruta R, Costantini V, Nascetti G (2006) Genetic diversity and phylogeography of the Apennine yellow-bellied toad Bombina pachypus, with implications for conservation. *Molecular Ecology*, **15**, 3741–3754.

Canestrelli D, Cimmaruta R, Nascetti G (2008) Population genetic structure

and diversity of the Apennine endemic stream frog, Rana italica – insights on the Pleistocene evolutionary history of the Italian peninsular biota. *Molecular Ecology*, **17**, 3856–3872.
Canestrelli D, Nascetti G (2008) Phylogeography of the pool frog Rana (Pelophylax) lessonae in the Italian peninsula and Sicily: Multiple refugia, glacial expansions and nuclear-mitochondrial discordance. *Journal of Biogeography*, **35**, 1923–1936.
Cano JM, Laurila A, Palo J, Merilä J (2004) Population differentiation in G matrix structure due to natural selection in Rana temporaria. *Evolution*, **58**, 2013–2020.
Catchen JM, Amores A, Hohenlohe P, Cresko W, Postlethwait JH (2011) Stacks: Building and Genotyping Loci De Novo From Short-Read Sequences. *G3:Genes, Genomes, Genetics*, **1**, 171–182.
Caye K, Deist TM, Martins H, Michel O, François O (2016) TESS3: Fast inference of spatial population structure and genome scans for selection. *Molecular Ecology Resources*, **16**, 540–548.
CH2014–Impacts (2014) *Toward quantitative scenarios of climate change Impacts in Switzerland*. OCCR, FOEN, MeteoSwiss, C2SM, Agroscope, and ProClim, Bern, Switzerland.
Clement M, Posada D, Crandall KA (2000) TCS : a computer program to estimate gene genealogies. *Molecular Ecology*, **9**, 1657–1659.
Cornetti L, Lemoine M, Hilfiker D *et al.* (2016) Higher genetic diversity on mountain tops: the role of historical and contemporary processes in shaping genetic variation in the bank vole. *Biological Journal of the Linnean Society*, **118**, 233–244.
Croteau MC, Davidson MA, Lean DRS, Trudeau VL (2008) Global Increases in Ultraviolet B Radiation : Potential Impacts on Amphibian Development and Metamorphosis. *Physiological and Biochemical Zoology*, **81**, 743–761.
Crottini A, Andreone F, Kosuch J *et al.* (2007) Fossorial but widespread: the phylogeography of the common spadefoot toad (Pelobates fuscus), and the role of the Po Valley as a major source of genetic variability. *Molecular Ecology*, **16**, 2734–2754.
Danecek P, Auton A, Abecasis G *et al.* (2011) The variant call format and VCFtools. *Bioinformatics*, **27**, 2156–2158.
Dansgaard W, Johnsen SJ, Clausen HB *et al.* (1993) Evidence for general instability of past climate from a 250-kyr ice-core record. *Nature*, **364**, 218–220.
Darnault R, Rolland Y, Braucher R *et al.* (2011) Timing of the last deglaciation revealed by receding glaciers at the Alpine-scale : impact on mountain geomorphology. *Quaternary Science Reviews*, 1–16.
Darriba D, Taboada GL, Doallo R, Posada D (2012) jModelTest 2: more models, new heuristics and parallel computing. *Nature Methods*, **9**, 772.
Debieu M, Tang C, Stich B *et al.* (2013) Co-Variation between seed dormancy, growth rate and flowering time changes with latitude in Arabidopsis thaliana. *PloS one*, **8**, 1–12.
Demesure B, Comps B, Petit RJ (1996) Chloroplast DNA Phylogeography of the Common Beech (Fagus sylvatica L.) in Europe. *Evolution*, **50**, 2515–2520.
Drummond AJ, Suchard MA, Xie D, Rambaut A (2012) Bayesian

Phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution*, **29**, 1969–1973.
Dudaniec RY, Spear SF, Richardson JS, Storfer A (2012) Current and historical drivers of landscape genetic structure differ in core and peripheral salamander populations. *PloS one*, **7**.
Eaton DAR (2014) PyRAD: Assembly of de novo RADseq loci for phylogenetic analyses. *Bioinformatics*, **30**, 1844–1849.
Ellis N, Smith SJ, Pitcher CR (2012) Gradient forests : calculating importance gradients on physical predictors. *Ecology*, **93**, 156–168.
Excoffier L, Lischer HEL (2010) An Integrated Software Package for Population Genetics Data Analysis. *Molecular Ecology Resources*, **10**, 564–567.
Fitzpatrick SW, Gerberich JC, Kronenberger JA, Angeloni LM, Funk WC (2015) Locally adapted traits maintained in the face of high gene flow. *Ecology Letters*, **18**, 37–47.
Fitzpatrick MC, Keller SR (2015) Ecological genomics meets community-level modelling of biodiversity: mapping the genomic landscape of current and future environmental adaptation. *Ecology Letters*, **18**, 1–16.
Forester BR, Jones MR, Joost S, Landguth EL, Lasky JR (2016) Detecting spatial genetic signatures of local adaptation in heterogeneous landscapes. *Molecular Ecology*, **25**, 104–120.
Fox J, Weisberg S (2011) *An {R} Companion to Applied Regression*. Thousand Oaks CA: Sage.
Freedman AH, Thomassen HA, Buermann W, Smith TB (2010) Genomic signals of diversification along ecological gradients in a tropical lizard. *Molecular ecology*, **19**, 3773–3788.
Frichot E, François O (2015) LEA : An R package for landscape and ecological association studies (B O'Meara, Ed,). *Methods in Ecology and Evolution*, **6**, 925–929.
Frichot E, Schoville SD, Bouchard G, François O (2013) Testing for associations between loci and environmental gradients using latent factor mixed models. *Molecular Biology and Evolution*, **30**, 1687–99.
Frichot E, Schoville S, de Villemereuil P, Gaggiotti OE, François O (2015) Detecting adaptive evolution based on association with ecological gradients: Orientation matters! *Heredity*, **115**, 22–28.
Garcia VOS, Ivy C, Fu J (2017) Syntopic frogs reveal different patterns of interaction with the landscape: A comparative landscape genetic study of Pelophylax nigromaculatus and Fejervarya limnocharis from central China. *Ecology and Evolution*, **7**, 9294–9306.
Gosner KL (1960) A Simplified Table for Staging Anuran Embryos and Larvae with Notes on Identification. *Herpetologica*, **16**, 183–190.
Goudet J (2005) Hierfstat, a package for R to compute and test hierarchical F-statistics. *Molecular Ecology Notes*, **5**, 184–186.
Gugerli F, Englisch T, Niklfeld H *et al.* (2008) Relationships among levels of biodiversity and the relevance of intraspecific diversity in conservation – a project synopsis. *Perspectives in Plant Ecology, Evolution and Systematics*, **10**, 259–281.
Günther T, Coop G (2013) Robust identification of local adaptation from allele frequencies. *Genetics*, **195**, 205–220.
Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor

and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series*, **41**, 95–98.
Harrisson KA, Amish SJ, Pavlova A *et al.* (2017) Signatures of polygenic adaptation associated with climate across the range of a threatened fish species with high genetic connectivity. *Molecular ecology*, **26**, 6253–6269.
Hecht BC, Matala AP, Hess JE, Narum SR (2015) Environmental adaptation in Chinook salmon (Oncorhynchus tshawytscha) throughout their North American range. *Molecular Ecology*, **24**, 5573–5595.
Heiri C, Bugmann H, Tinner W, Heiri O, Lischke H (2006) A model-based reconstruction of Holocene treeline dynamics in the Central Swiss Alps. *Journal of Ecology*, **94**, 206–216.
Hermisson J, Pennings PS (2005) Soft sweeps: molecular population genetics of adaptation from standing genetic variation. *Genetics*, **169**, 2335–52.
Hewitt GM (1996) Some genetic consequences of ice ages, and their role in divergence and speciation. *Biological Journal of the Linnean Society*, **58**, 247–276.
Hewitt G (1999) Post-glacial re-colonization of European biota. *Biological Journal of the Linnean Society*, **68**, 87–112.
Hewitt G (2000) The genetic legacy of the Quaternary ice ages. *Nature*, **405**, 907–913.
Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A (2005) Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, **25**, 1965–1978.
Hitchings SP, Beebee TJC (1997) Genetic substructuring as a result of barriers to gene flow in urban Rana temporaria (common frog) populations: implications for biodiversity conservation. *Heredity*, **79**, 117–127.
Hjernquist MB, Soderman F, Jonsson KI *et al.* (2012) Seasonality determines patterns of growth and age structure over a geographic gradient in an ectothermic vertebrate. *Oecologia*, **170**, 641–649.
Hoban S, Kelley JL, Lotterhos KE *et al.* (2016) Finding the Genomic Basis of Local Adaptation: Pitfalls, Practical Solutions, and Future Directions. *The American Naturalist*, **188**, 000–000.
Hoffmann A a, Sgrò CM (2011) Climate change and evolutionary adaptation. *Nature*, **470**, 479–85.
Huelsenbeck JP, Ronquist F (2001) MRBAYES : Bayesian inference of phylogenetic trees. *Bioinformatics Application Note*, **17**, 754–755.
Ilut DC, Nydam ML, Hare MP (2014) Defining loci in restriction-based reduced representation genomic data from nonmodel species: Sources of bias and diagnostics for optimal clustering. *BioMed Research International*, **2014**, 9 pages.
Jombart T (2008) Adegenet: A R package for the multivariate analysis of genetic markers. *Bioinformatics*, **24**, 1403–1405.
Jombart T, Ahmed I (2011) adegenet 1.3-1: New tools for the analysis of genome-wide SNP data. *Bioinformatics*, **27**, 3070–3071.
Jones MR, Forester BR, Teufel AI *et al.* (2013) Integrating landscape genomics and spatially explicit approaches to detect loci under selection in clinal populations. *Evolution*, **67**, 3455–68.
Kawecki TJ, Ebert D (2004) Conceptual issues in local adaptation. *Ecology*

Letters, **7**, 1225–1241.
Lasky JR, Des Marais DL, McKay J *et al.* (2012) Characterizing genomic variation of Arabidopsis thaliana: the roles of geography and climate. *Molecular Ecology*, **21**, 5512–5529.
Laugen AT, Kruuk LEB, Laurila A *et al.* (2005a) Quantitative genetics of larval life-history traits in Rana temporaria in different environmental conditions. *Genetics Research*, **86**, 161–170.
Laugen AT, Laurila A, Jönsson KI, Söderman F, Merilä J (2005b) Do common frogs (*Rana temporaria*) follow Bergmann's rule? *Evolutionary Ecology Research*, **7**, 717–731.
Laugen AT, Laurila A, Merilä J (2002) Maternal and genetic contributions to geographical variation in Rana temporaria larval life-history traits. *Biological Journal fo the Linnean Society*, **76**, 61–70.
Laugen AT, Laurila A, Merilä J (2003a) Latitudinal and temperature-dependent variation in embryonic development and growth in Rana temporaria. *Oecologia*, **135**, 548–554.
Laugen AT, Laurila A, Rasanen K, Merilä J (2003b) Latitudinal countergradient variation in the common frog (Rana temporaria) development rates – evidence for local adaptation. *Journal of Evolutionary Biology*, **16**, 996–1005.
Laurila A, Karttunen S, Meril J (2002) Adaptive Phenotypic Plasticity and Genetics of Larval Life Histories in Two Rana Temporaria Populations. *Evolution*, **56**, 617–627.
Laurila A, Pakkasmaa S, Merila J (2001) Influence of Seasonal Time Constraints on Growth and Development of Common Frog Tadpoles: A Photoperiod Experiment. *Oikos*, **95**, 451–460.
Legendre P, Fortin M-J (2010) Comparison of the Mantel test and alternative approaches for detecting complex multivariate relationships in the spatial analysis of genetic data. *Molecular Ecology Resources*, **10**, 831–844.
Legendre P, Legendre LF (2012) *Numerical ecology*. Elsevier.
Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics Application Note*, **25**, 1451–1452.
Lindgren B, Laurila A (2005) Proximate causes of adaptive growth rates: growth efficiency variation among latitudinal populations of Rana temporaria. *Journal of Evolutionary Biology*, **18**, 820–828.
Lindgren B, Laurila A (2009) Physiological variation along a geographical gradient: is growth rate correlated with routine metabolic rate in Rana temporaria tadpoles? *Biological Journal of the Linnean Society*, **98**, 217–224.
Loman J, Claesson D (2003) Plastic response to pond drying in tadpoles Rana temporaria: tests of cost models. *Evolutionary Ecology Research*, **5**, 179–194.
Lugon-Moulin N, Hausser J (2002) Phylogeographical structure, postglacial recolonization and barriers to gene flow in the distinctive Valais chromosome race of the common shrew (Sorex araneus). *Molecular Ecology*, **11**, 785–794.
Luikart G, England PR, Tallmon D, Jordan S, Taberlet P (2003) The power and promise of population genomics: from genotyping to genome typing. *Nature Reviews Genetics*, **4**, 981–94.

Luu K, Bazin E, Blum MGB (2016) pcadapt: An R package to perform genome scans for selection based on principal component analysis. *Molecular Ecology Resources*, **33**, 67–77.
Manel S, Perrier C, Pratlong M *et al.* (2016) Genomic resources and their influence on the detection of the signal of positive selection in genome scans. *Molecular ecology*, **25**, 170–184.
Manel S, Poncet BN, Legendre P, Gugerli F, Holderegger R (2010) Common factors drive adaptive genetic variation at different spatial scales in Arabis alpina. *Molecular ecology*, **19**, 3824–35.
Marquis O, Miaud C (2008) Variation in UV sensitivity among common frog Rana temporaria populations along an altitudinal gradient. *Zoology*, **111**, 309–317.
Marquis O, Miaud C, Lena J-P (2008) Developmental responses to UV-B radiation in common frog Rana temporaria embryos from along an altitudinal gradient. *Population Ecology*, **50**, 123–130.
Mátyás G, Sperisen C (2001) Chloroplast DNA polymorphisms provide evidence for postglacial re-colonisation of oaks (Quercus spp.) across the Swiss Alps. *Theoretical and Applied Genetics*, **102**, 12–20.
McCain CM, Colwell RK (2011) Assessing the threat to montane biodiversity from discordant shifts in temperature and precipitation in a changing climate. *Ecology Letters*, **14**, 1236–1245.
Merila J, Laurila A, Laugen AT, Rasanen K, Pahkala M (2000) Plasticity in age and size at metamorphosis in Rana temporaria: comparison of high and low latitude populations. *Ecography*, **23**, 457–465.
Merilä J, Laurila A, Laugen AT, Räsänen K, Pahkala M (2000) Plasticity in Age and Size at Metamorphosis in Rana Temporaria: Comparison of High and Low Latitude Populations. *Ecography*, **23**, 457–465.
Messer PW, Petrov DA (2013) Population genomics of rapid adaptation by soft selective sweeps. *Trends in Ecology and Evolution*, **28**, 659–669.
Miaud C, Guyetant R, Elmberg J (1999) Variations in life-history traits in the common frog Rana temporaria (Amphibia: Anura): a literature review and new data from the French Alps. *Journal of Zoology*, **249**, 61–73.
Miaud C, Merilä J (2001) Local adaptation or environmental induction? Causes of population differentiation in alpine amphibians. *Biota*, **2**, 31–50.
Muir AP, Biek R, Mable BK (2014a) Behavioural and physiological adaptations to low-temperature environments in the common frog, Rana temporaria. *BMC Evolutionary Biology*, **14**, 1471–2148.
Muir A, Piek R, Thomas R, Mable B (2014b) Local adaptation with high gene flow: temperature parameters drive adaptation to altitude in the common frog (Rana temporaria). *Molecular ecology*, **23**, 561–574.
Nychka D, Furrer R, Paige J, Sain S (2015) fields: Tools for spatial data.
Oksanen J, Blanchet FG, Kindt R *et al.* (2015) Vegan: community ecology package. R package vegan, vers. 2.2-1.
Pahkala M, Laurila A, Meril J (2002) Effects of ultraviolet-B radiation on common frog Rana temporaria embryos from along a latitudinal gradient. *Oecologia*, **133**, 458–465.
Pahkala M, Laurila A, Merila J (2000) Ambient Ultraviolet-B radiation reduces hatchling size in the common frog Rana temporaria. *Ecography*, **23**, 531–538.

Pahkala M, Laurila A, Merila J (2001) Carry-over effects of ultraviolet-B radiation on larval tness in Rana temporaria. *Proceedings of the Royal Society of London B: Biological Sciences*, **268**, 1699–1706.

Pahkala M, Merila J, Ots I, Laurila A (2003) Effects of ultraviolet-B radiation on metamorphic traits in the common frog Rana temporaria. *Journal of the Zoological Society of London*, **259**, 57–62.

Palo JU, O'Hara RB, Laugen AT *et al.* (2003a) Latitudinal divergence of common frog (Rana temporaria) life history traits by natural selection: evidence from a comparison of molecular and quantitative genetic data. *Molecular ecology*, **12**, 1963–1978.

Palo JU, O'Hara RB, Laugen AT *et al.* (2003b) Latitudinal divergence of the common frog (Rana temporaria) life history traits by natural selection: evidence from a comparison of molecular quantitative genetic data. *Molecular Ecology*, **12**, 1963–1978.

Palo JU, Schmeller DS, Laurila A *et al.* (2004) High degree of population subdivision in a widespread amphibian. *Molecular Ecology*, **13**, 2631–2644.

Parisod C (2008) Postglacial recolonisation of plants in the western Alps of Switzerland. *Botanica Helvetica*, **118**, 1–12.

Parmesan C (2006) Ecological and Evolutionary Responses to Recent Climate Change. *Annual Review of Ecology, Evolution, and Systematics*, **37**, 637–669.

Parmesan C, Yohe G (2003) A globally coherent fingerprint of climate change impacts across natural systems. *Nature*, **421**, 37–42.

Pennings PS, Hermisson J (2006) Soft sweeps II - Molecular population genetics of adaptation from recurrent mutation or migration. *Molecular Biology and Evolution*, **23**, 1076–1084.

Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE (2012) Double Digest RADseq: An Inexpensive Method for De Novo SNP Discovery and Genotyping in Model and Non-Model Species. *PloS one*, **7**.

Polechová J, Barton NH (2015) Limits to adaptation along environmental gradients. *Proceedings of the National Academy of Sciences of the United States of America*, **112**, 6401–6406.

Poncet BN, Herrmann D, Gugerli F *et al.* (2010) Tracking genes of ecological relevance using a genome scan in two independent regional population samples of Arabis alpina. *Molecular ecology*, **19**, 2896–907.

Pritchard JK, Di Rienzo A (2010) Adaptation - not by sweeps alone. *Nature Reviews Genetics*, **11**, 665–7.

Purcell S, Neale B, Todd-Brown K *et al.* (2007) PLINK: A tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics*, **81**, 559–575.

Raj A, Stephens M, Pritchard JK (2014) FastSTRUCTURE: Variational inference of population structure in large SNP data sets. *Genetics*, **197**, 573–589.

Rehm EM, Olivas P, Stroud J, Feeley KJ (2015) Losing your edge: climate change and the conservation value of range-edge populations. *Ecology and evolution*, **5**, 4315–4326.

Rellstab C, Fischer MC, Zoller S *et al.* (2016) Local adaptation (mostly) remains local: reassessing environmental associations of climate-related candidate SNPs in Arabidopsis halleri. *Heredity*, 1–9.

Rellstab C, Gugerli F, Eckert AJ, Hancock AM, Holderegger R (2015) A practical guide to environmental association analysis in landscape genomics. *Molecular Ecology*, **24**, 4348–70.
Rigling A, Bigler C, Eilmann B *et al.* (2013) Driving factors of a vegetation shift from Scots pine to pubescent oak in dry Alpine forests. *Global Change Biology*, **19**, 229–240.
Rodrigues N, Betto-Colliard C, Jourdan-Pineau H, Perrin N (2013) Within-population polymorphism of sex-determination systems in the common frog (Rana temporaria). *Journal of evolutionary biology*, **26**, 1569–1577.
Roesti M, Salzburger W, Berner D (2012) Uninformative polymorphisms bias genome scans for signatures of selection. *BMC Evolutionary Biology*, **12**, 94.
Roff DA (1996) The evolution of threshold traits in animals. *The Quarterly Review of Biology*, **71**, 3–35.
Rogers AR, Harpending H (1992) Population Growth Makes Waves in the Distribution of Pairwise Genetic Differences. *Molecular biology and evolution*, **9**, 552–569.
Rogivue A, Graf R, Parisod C, Holderegger R, Gugerli F (2018) The phylogeographic structure of Arabis alpina in the Alps shows consistent patterns across different types of molecular markers and geographic scales. *Alpine Botany*, **0**, 0.
Rousset F (1997) Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. *Genetics*, **145**, 1219–1228.
Roy K, Valentine JW, Jablonski D, Kidwell SM (1996) Scales of climatic variability and time averaging in Pleistocene biotas: implications for ecology and evolution. *Trends in Ecology & Evolution*, **11**, 458–463.
Schmeller DS, Palo JU, Merilä J (2008) A contact zone between two distinct Rana temporaria lineages in northern Germany. *Alytes*, **25**, 93–98.
Schneider S, Excoffier L (1999) Estimation of Past Demographic Parameters From the Distribution of Pairwise Differences When the Mutation Rates Vary Among Sites: Application to Human Mitochondrial DNA. *Genetics*, **152**, 1079–1089.
Schweizer RM, VonHoldt BM, Harrigan R *et al.* (2016) Genetic subdivision and candidate genes under selection in North American grey wolves. *Molecular Ecology*, **25**, 380–402.
Sillero N, Campos J, Bonardi A *et al.* (2014) Updated distribution and biogeography of amphibians and reptiles of Europe. *Amphibia-Reptilia*, **35**, 1–31.
Sork VL, Aitken SN, Dyer RJ *et al.* (2013) Putting the landscape into the genomics of trees: approaches for understanding local adaptation and population responses to changing climate. *Tree Genetics & Genomes*, **9**, 901–911.
Ståhlberg F, Olsson M, Uller T (2001) Population divergence of developmental thermal optima in Swedish common frogs, Rana temporaria. *Journal of Evolutionary Biology*, **14**, 755–762.
Stamatakis A (2006) RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics Application Note*, **22**, 2688–2690.
Stapley J, Reger J, Feulner PGD *et al.* (2010) Adaptation genomics: the next generation. *Trends in ecology & evolution*, **25**, 705–12.

Stefani F, Gentilli A, Sacchi R *et al.* (2012) Refugia within refugia as a key to disentangle the genetic pattern of a highly variable species: The case of Rana temporaria Linnaeus, 1758 (Anura, Ranidae). *Molecular Phylogenetics and Evolution*, **65**, 718–726.

Stehlik I, Blattner FR, Holderegger R, Bachmann K (2002) Nunatak survival of the high Alpine plant Eritrichium nanum (L.) Gaudin in the central Alps during the ice ages. *Molecular Ecology*, **11**, 2027–2036.

Taberlet P, Fumagalli L, Wust-Saucy A-G, Cossons J-F (1998) Comparative phylogeography and postglacial colonization routes in Europe. *Molecular Ecolgy*, **7**, 453–464.

Teacher AGF, Garner TWJ, Nichols RA (2009) European phylogeography of the common frog (Rana temporaria): routes of postglacial colonization into the British Isles, and evidence for an Irish glacial refugium. *Heredity*, **102**, 490–6.

Templeton AR, Crandall KA, Sing CF (1992) Cladistic Analysis of Phenotypic Associations With Haplotypes Inferred From Restriction Endonuclease Mapping and DNA Sequence Data. III. Cladogram Estimation. *Genetics*, **132**, 619–633.

Thomassen H a, Cheviron Z a, Freedman AH *et al.* (2010) Spatial modelling and landscape-level approaches for visualizing intra-specific variation. *Molecular Ecology*, **19**, 3532–48.

Tinner W, Theurillat J (2003) Uppermost limit, extent, and fluctuations of the timberline and treeline ecocline in the Swiss Central Alps during the past 11,500 years. *Arctic, Antarctic, and Alpine Research*, **35**, 158–169.

Veith M, Kosuch J, Vences M (2003) Climatic oscillations triggered post-Messinian speciation of Western Palearctic brown frogs (Amphibia, Ranidae). *Molecular phylogenetics and evolution*, **26**, 310–327.

Veith M, Vences M, Vieites DR, Nieto-roman S, Palanca A (2002) Genetic differentiation and population structure within Spanish common frogs (Rana temporaria complex; Ranidae, Amphibia). *Folia Zoologica*, **51**, 307–318.

Vences M, Hauswaldt JS, Steinfartz S *et al.* (2013) Radically different phylogeographies and patterns of genetic variation in two European brown frogs, genus Rana. *Molecular phylogenetics and evolution*, **68**, 657–70.

Vergeer P, Kunin WE (2013) Adaptation at range margins: common garden trials and the performance of Arabidopsis lyrata across its northwestern European range. *New Phytologist*, **197**, 989–1001.

Vitti JJ, Grossman SR, Sabeti PC (2013) Detecting Natural Selection in Genomic Data. *Annu. Rev. Genet*, **47**, 97–120.

Wang IJ (2012) Environmental and topographic variables shape genetic structure and effective population sizes in the endagered Yosemite toad. *Diversity and Distributions*, **18**, 1033–1041.

Wang IJ, Glor RE, Losos JB (2013) Quantifying the roles of ecology and geography in spatial genetic divergence. *Ecology letters*, **16**, 175–82.

Weir BS, Cockerham CC (1984) Estimating F-Statistics for the Analysis of Population Structure. *Evolution*, **38**, 1358–1370.

Wigginton JE, Cutler DJ, Abecasis GR (2005) A note on exact tests of Hardy-Weinberg equilibrium. *American Journal of Human Genetics*, **76**, 887–93.

Willis KJ, Whittaker RJ (2008) The refugial debate. *Science*, **287**, 1406–1407.

Yannic G, Basset P, Hausser J (2008) Phylogeography and recolonization of the Swiss Alps by the Valais shrew (Sorex antinorii), inferred with autosomal and sex-specific markers. *Molecular Ecology*, **17**, 4118–4133.
Yannic G, Pellissier L, Dubey S *et al.* (2012) Multiple refugia and barriers explain the phylogeography of the Valais shrew, Sorex antinorii (Mammalia: Soricomorpha). *Biological Journal of the Linnean Society*, **105**, 864–880.
Yeaman S (2015) Local Adaptation by Alleles of Small Effect. *The American Naturalist*, **186**, S74–S89.

Appendix Chapter 2

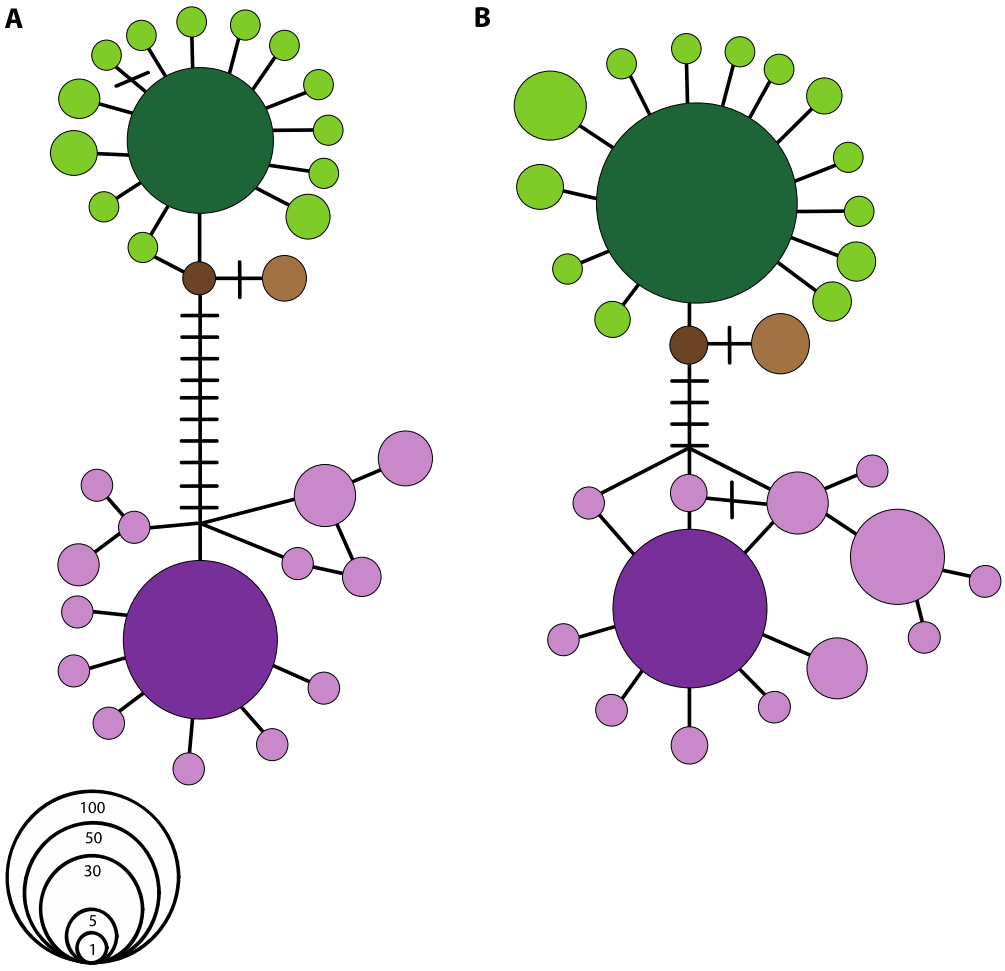


Figure S2.1 Haplotype network of individuals sampled in Switzerland only. a) 448bp cytb sequenced for 368 individuals from 72 populations, b) 1076bp of mitochondrial DNA (448bp cytb and 628bp COX1) sequenced for 44 populations (151 individuals) from Switzerland. Circles represent haplotypes, and are scaled by sample size. Colours correspond to those in Fig. 2.1. A solid line represents a single mutational step linking two haplotypes, with cross hatches added for additional steps.

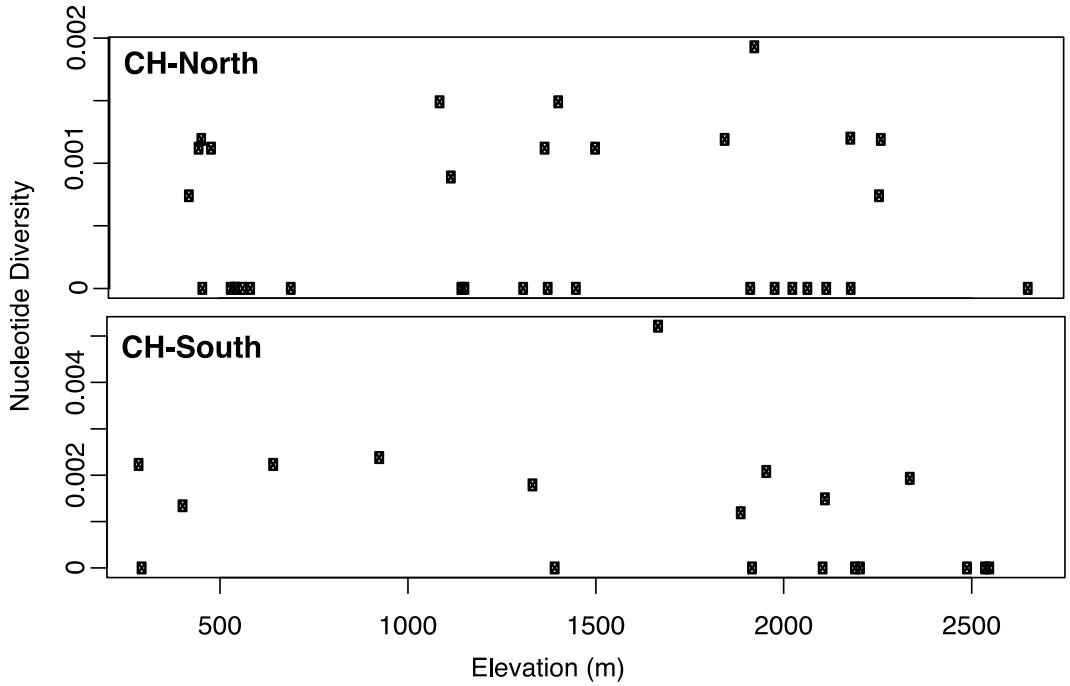


Figure S2.2 Nucleotide diversity plotted across elevation for CH-North (top) and CH-South (bottom) from Switzerland. We excluded populations where the haplogroups co-occur.

Table S2.1 Populations sampled and sequenced at cytochrome b. N= number of samples sequenced, Haplotypes= number of haplotypes found in the population, Hd (SD)= Haplotype diversity (standard deviation), nd (SD) = nucleotide diversity (standard deviation).

Population	Lat.	Long.	Elevation	N	length (bp)	Haplotypes	Hd (SD)	nd (SD)
agra	46.03	8.90	930	6	448	3	0.8 (0.122)	0.00238 (0.0005)
alpl	46.94	8.66	1506	4	448	2	0.5 (0.27)	0.00112 (0.00059)
apla	46.79	9.50	2118	6	448	3	0.6 (0.22)	0.0091 (0.00297)
arce	46.16	8.75	408	6	448	3	0.73 (0.16)	0.0064 (0.00251)
bach	46.67	8.03	2265	6	448	2	0.53 (0.17)	0.00119 (0.00038)
bela	46.38	7.98	2179	6	448	3	0.73 (0.16)	0.00699 (0.00288)
bend	46.49	6.88	590	6	448	1	0 (0)	0 (0)
berm	47.02	7.52	550	6	448	1	0 (0)	0 (0)
bide	47.14	9.40	1983	6	448	1	0 (0)	0 (0)
birk	47.30	8.81	550	4	448	1	0 (0)	0 (0)
bnnp	46.41	10.03	2342	6	448	3	0.73 (0.16)	0.00193 (0.00055)
buel	46.86	9.73	2260	6	448	2	0.33 (0.22)	0.00074 (0.00048)
cava	46.36	9.03	2003	6	448	2	0.33 (0.22)	0.00521 (0.00336)
csan	46.32	7.30	2110	6	448	1	0 (0)	0 (0)
egel	47.25	9.50	445	6	448	2	0.33 (0.22)	0.00595 (0.00384)
fada	46.98	9.60	1093	3	448	2	0.67 (0.31)	0.00149 (0.00070)
fess	47.02	9.14	2184	4	448	2	0.5 (0.27)	0.0012 (0.00059)
flue	46.75	9.95	2388	3	448	3	0.83 (0.05)	0.0115 (0.00335)
forn	46.48	8.58	2089	5	448	3	0.7 (0.22)	0.00982 (0.00295)
full	46.17	7.10	2074	5	448	3	0.7 (0.22)	0.00759 (0.00333)
fuor	46.44	9.83	2494	6	448	1	0 (0)	0 (0)
gdwe	46.64	8.07	1380	6	448	1	0 (0)	0 (0)

gola	46.10	8.97	975	6	448	5	0.93 (0.12)	0.00774 (0.00282)
gott	46.59	8.56	2116	6	448	2	0.33 (0.22)	0.00149 (0.00096)
grma	46.24	7.01	425	6	448	3	0.6 (0.22)	0.00714 (0.00356)
grsh	46.66	8.10	1929	6	448	3	0.73 (0.16)	0.00193 (0.00055)
gruu	46.86	9.79	2120	4	448	1	0 (0)	0 (0)
hdns	46.85	9.76	1918	4	448	1	0 (0)	0 (0)
jagg	46.74	8.06	570	4	448	1	0 (0)	0 (0)
kand	46.63	7.69	698	4	448	1	0 (0)	0 (0)
kebe	47.54	8.78	453	4	448	2	0.5 (0.27)	0.00112 (0.00059)
lens	46.29	7.46	1338	5	448	2	0.4 (0.24)	0.00179 (0.00106)
lucm	46.56	8.80	1922	6	448	1	0 (0)	0 (0)
magn	46.43	8.68	1840	4	448	2	0.5 (0.27)	0.0067 (0.00355)
mart	46.03	8.94	407	5	448	2	0.6 (0.18)	0.00134 (0.00039)
mctn	46.48	9.72	2542	5	448	1	0 (0)	0 (0)
mgns	46.25	6.85	1372	4	448	2	0.5 (0.27)	0.00112 (0.00059)
moau	46.34	6.79	1850	6	448	2	0.53 (0.17)	0.00119 (0.00038)
moir	46.10	7.57	2553	4	448	1	0 (0)	0 (0)
muet	47.45	8.61	460	6	448	2	0.53 (0.17)	0.00119 (0.00038)
munt	46.73	9.44	648	2	448	2	1 (0.5)	0.00223 (0.00112)
oalp	46.65	8.64	1960	6	448	3	0.73 (0.16)	0.00208 (0.00058)
otte	46.73	7.36	1455	6	448	1	0 (0)	0 (0)
petl	45.89	7.15	2655	4	448	1	0 (0)	0 (0)
pizo	46.98	9.42	2209	4	448	1	0 (0)	0 (0)
pozz	46.35	8.96	290	4	448	3	0.83 (0.22)	0.00223 (0.00076)
prad	46.78	9.53	1449	5	448	4	0.9 (0.16)	0.01161 (0.00307)
roes	46.90	7.20	538	6	448	1	0 (0)	0 (0)
rose	46.14	7.05	450	4	448	3	0.83 (0.22)	0.01042 (0.00367)

rotc	47.01	9.31	2185	4	448	1	0 (0)	0 (0)
rusc	46.37	7.24	1315	4	448	1	0 (0)	0 (0)
sali	46.26	8.69	331	4	448	3	0.83 (0.22)	0.00967 (0.00383)
saxm	47.07	9.38	463	9	448	1	0 (0)	0 (0)
scai	45.98	8.93	298	2	448	1	0 (0)	0 (0)
seeo	47.05	9.58	2030	6	448	1	0 (0)	0 (0)
seji	46.80	9.73	2090	6	448	2	0.83 (0.22)	0.00595 (0.00384)
shwe	47.19	9.33	1159	6	448	1	0 (0)	0 (0)
siec	46.99	9.55	517	5	448	2	0.6 (0.18)	0.00938 (0.00274)
star	46.27	8.77	1892	6	448	2	0.53 (0.17)	0.00119 (0.00038)
stba	46.49	9.17	2059	6	448	2	0.33 (0.22)	0.00521 (0.00336)
stir	46.53	7.55	2070	4	448	1	0 (0)	0 (0)
stls	46.97	9.75	1672	6	448	2	0.33 (0.22)	0.00521 (0.00336)
tana	46.35	6.84	1408	6	448	3	0.6 (0.22)	0.00149 (0.00062)
trim	46.90	9.54	545	5	448	2	0.4 (0.24)	0.00625 (0.00371)
trpa	46.29	7.28	2196	4	448	1	0 (0)	0 (0)
tsee	46.55	7.75	1151	6	448	1	0 (0)	0 (0)
vilt	47.03	9.45	486	4	448	2	0.5 (0.27)	0.00112 (0.00059)
vora	47.16	9.38	1123	5	448	2	0.4 (0.24)	0.00089 (0.00053)
wale	47.12	9.10	427	6	448	2	0.33 (0.22)	0.00074 (0.00048)
wdji	46.81	9.72	1638	6	448	3	0.8 (0.12)	0.01071 (0.00303)
wise	47.19	9.48	445	5	448	4	0.9 (0.16)	0.01027 (0.00247)
zeni	46.52	8.89	1397	5	448	1	0 (0)	0 (0)

Table S2.2 Populations sampled and sequenced at COX1. N = number of samples sequenced, Haplotypes = number of haplotypes found in the population, Hd (SD) = Haplotype diversity (standard deviation), nd (SD) = nucleotide diversity (standard deviation).

Population	N	length (bp)	nr haplotypes	Hd (SD)	nucleotide diversity (SD)
alpl	4	628	1	0 (0)	0 (0)
arce	3	628	1	0 (0)	0 (0)
bach	4	628	1	0 (0)	0 (0)
bela	3	628	1	0 (0)	0 (0)
bide	3	628	2	0.67 (0.31)	0.00106 (0.00050)
birk	4	628	1	0 (0)	0 (0)
bnnp	3	628	2	0.67 (0.31)	0.00106 (0.00050)
cava	3	628	2	0.67 (0.31)	0.00637 (0.003)
egel	3	628	2	0.67 (0.31)	0.00106 (0.00050)
fada	3	628	1	0 (0)	0 (0)
flue	4	628	3	0.83 (0.22)	0.00717 (0.00207)
forn	4	628	3	0.83 (0.22)	0.00717 (0.00207)
fuor	3	628	1	0 (0)	0 (0)
gott	4	628	2	0.5 (0.27)	0.0008 (0.00042)
grma	4	628	2	0.5 (0.27)	0.0008 (0.00042)
grsh	4	628	1	0 (0)	0 (0)
gruu	4	628	2	0.5 (0.27)	0.0008 (0.00042)
hdns	4	628	2	0.5 (0.27)	0.0008 (0.00042)
jagg	4	628	1	0 (0)	0 (0)
kand	4	628	2	0.5 (0.27)	0.00159 (0.00084)
kebe	4	628	1	0 (0)	0 (0)
magn	3	628	2	0.67 (0.31)	0.00743 (0.0035)

mart	4	628	2	0.67 (0.2)	0.00212 (0.00065)
mctn	4	628	1	0 (0)	0 (0)
mgns	4	628	1	0 (0)	0 (0)
moir	4	628	1	0 (0)	0 (0)
petl	4	628	1	0 (0)	0 (0)
pizo	4	628	1	0 (0)	0 (0)
pozz	4	628	2	0.5 (0.27)	0.00159 (0.00084)
prad	4	628	2	0.5 (0.27)	0.00478 (0.00253)
rotc	4	628	1	0 (0)	0 (0)
rusc	4	628	1	0 (0)	0 (0)
seeo	4	628	2	0.67 (0.2)	0.00106 (0.00033)
shwe	4	628	1	0 (0)	0 (0)
siec	1	628	1	0 (0)	0 (0)
star	1	628	1	0 (0)	0 (0)
stba	3	628	3	1.0 (0.27)	0.00743 (0.00304)
stir	2	628	1	0 (0)	0 (0)
stls	1	628	1	0 (0)	0 (0)
tana	3	628	1	0 (0)	0 (0)
trpa	4	628	1	0 (0)	0 (0)
vora	2	628	1	0 (0)	0 (0)
wdji	4	628	2	0.5 (0.27)	0.00478 (0.00253)
zeni	3	628	1	0 (0)	0 (0)

Population structure reflects colonisation and elevation does not impede gene flow in the European common frog in the Swiss Alps

Mountains present a complex topography and heterogeneous environments and often result in complex population structure across its range. The European Alps are particularly interesting because in addition to the steep terrain they present a large environmental gradient across a relatively small geographic distance. The European common frog, *R. temporaria*, occurs over ~2000 m elevational gradient in the Alps, and there is evidence for adaptive divergence between high and low elevations. However, there's been no large-scale assessment of population structure and connectivity between populations within Switzerland or across elevation. My aim here was to determine how the Alps have shaped population structure of *R. temporaria* in Switzerland. Previous phylogeographic analyses identified two diverged mitochondrial lineages that occur roughly north and south of the alpine ridge bisecting Switzerland, with a contact zone in eastern and western Switzerland. Population structure analysis, based on genome-wide markers generated for 82 populations throughout their range in Switzerland, reflected the colonisation history. I find two geographic groups distributed roughly north and south of the bisecting alpine ridge. In addition I find that southern Switzerland can be divided into a south-western and south-eastern cluster that were likely colonised independently from different glacial refugia in Italy. An extensive contact zone between the northern and southeastern groups occurs in eastern Switzerland. I also find significant isolation by distance throughout Switzerland, but no significant effect of elevation on gene flow. However, elevation and geographic distance describe the same proportion of genetic variation. These results imply adaptive divergence across elevation despite high levels of gene flow.

Introduction

The origin and maintenance of genetic diversity is a central problem in population genetics partly because genetic diversity is so important for evolutionary resilience, adaptive potential, and the ability to colonise new habitats. Genetic diversity often varies across a species' range, and this implies that geographic context influences the historical and demographic processes that shape the genetic composition of populations. Montane habitats are particularly interesting to study in this context, since species are expected to shift their ranges upwards in elevation in response to climate change (Parmesan & Yohe 2003). High elevation populations have been highlighted as harboring genetic variation that is potentially important for adaptation to extreme environments (Rehm *et al.* 2015). Differences in genetic diversity within and between elevational gradients may be important reservoirs of genetic diversity. Thus how genetic variation is distributed across the range and how much gene flow there is between populations has important implications for the maintenance of genetic diversity.

Predictions about the distribution of genetic variation across elevational gradients are based on characteristics typical of populations inhabiting mountainous regions. For example, the environmental gradient is relatively continuous, colonisation is expected to proceed from low to high elevation, and populations at high elevation are assumed to be at or near their range limit. Thus we expect that genetic diversity will decrease with elevation. Moreover, the steepness of the cline in environmental conditions could influence the extent of contemporary gene flow across the gradient and (Polechová & Barton 2015). For example, an increase in genetic variation due to increased dispersal (Cornetti *et al.* 2016), or an increase at mid-elevations due to source-sink dynamics across the gradient.

The European Alps present a particularly interesting system to study because they have a complex colonisation history leading to isolation between different regions in Switzerland. Gene flow within and between the gradients can be challenging across the steep mountain ridges and topographic complexity. Thus high elevation populations can be rather

isolated, and the amount of intra- and inter-gradient variation is expected to be high.

Here we focus on a widespread amphibian species, *Rana temporaria*, to characterise the effects of colonisation history, isolation by distance, and isolation by elevation in shaping the geographic distribution of genetic variation in the Alps. *R. temporaria* presents a good system for several reasons. Firstly, they are widespread throughout Europe and have a broad elevational range compared to most plants and animals (up to 2500 m in the Alps). Secondly, they are well monitored and there is a good estimate of their elevational range limit in the Alps. Thirdly, population structure had been studied at low elevations in Switzerland and more broadly across Europe (Palo *et al.* 2004; Buskirk 2012; Muir *et al.* 2014). Finally, there is evidence of adaptation across elevation with divergence in several larval life history traits (Miaud *et al.* 1999; Miaud & Merilä 2001; Bonin *et al.* 2006; Muir *et al.* 2014b). Disentangling the effects of these different processes and characterising the range-wide variation in genetic diversity is important for predicting resilience across a species range.

We aim to investigate how colonisation history, isolation by distance, and isolation by elevation have shaped the geographic distribution of genetic diversity of *R. temporaria* in the Alps. We approach this question by describing genetic diversity and genetic distance in mitochondrial and genome-wide markers across different geographic regions in Switzerland. Our goals were to 1) determine whether population structure based on nuclear markers reflects the colonisation history determined by mitochondrial markers, 2) determine how geographic distance and elevation contribute to genetic distance between populations, and 3) determine how genetic diversity is distributed across the elevational gradient.

Materials and Methods

Sample Collection

To investigate the geographic distribution of genomic variation in Switzerland, we sampled 82 populations across a ~2300 m gradient of elevation in the Swiss Alps during the 2013 breeding season (Table S3.1). Approximately 10 eggs from each of 20-30 freshly laid clutches were collected from each site. The eggs were transported to the University of Zürich, where they were hatched in separate water-filled containers in a water bath. The tadpoles were euthanised and stored in ethanol when they reached stage 36 (Gosner, 1960). Each clutch is assumed to represent a single family; thus one tadpole per clutch was used for all work presented here.

Data generation

DNA extractions - Total DNA was extracted from tadpole tails from one individual per clutch (henceforth family). Tissue samples were digested overnight in 10% Proteinase K solution at 56 °C and extracted using the Qiagen Biosprint 96 DNA blood kit (Qiagen, CA, USA) and Biosprint 96 workstation (QIAgen, Venlo, Netherlands). DNA was eluted in 100 - 200 µL buffer AE (QIAgen).

Mitochondrial DNA - We partially sequenced the mitochondrial *cytb* gene for 368 individuals from 72 populations. We generated a 448 bp fragment using *Rana-cytb-F2* and *Rana-cytb-R2* primers (Vences *et al.* 2013). PCR cycling conditions are described elsewhere (Chapter 1).

ddRAD sequencing, de novo assembly, and variant calling - Genome-wide markers were established with a modified version of the double digest restriction-site-associated DNA protocol (Peterson *et al.* 2012). Using *EcoRI* and *MseI* restriction enzymes, we constructed libraries comprising 48

individually barcoded samples each for single-end sequencing (125bp) on an Illumina HiSeq 2500 v4 at the Functional Genomics Center, University of Zurich. To ensure even representation of samples in a library, we normalised samples after individual barcodes were added and prior to pooling based on Qubit dsDNA assays (Thermo Fisher Scientific). A total of 22 HiSeq libraries were sequenced, three of which were re-sequenced due to low read counts in the first run. Raw sequence reads were demultiplexed using process_radtags (Catchen *et al.* 2011) based on the 5-bp unique barcode + the 4-bp restriction enzyme recognition site, with one mismatch allowed. For the repeated libraries we concatenated the demultiplexed data from both sequencing runs for each individual. We used Trimmomatic v0.33 (Bolger *et al.* 2014) to remove adapter and Illumina-specific sequences from all individuals before further processing.

We used pyRAD (Eaton 2014) for *de novo* assembly and variant calling. This pipeline clusters sequences for each individual separately based on a user specified clustering threshold; that is the percentage of base pairs that should align between sequences for a cluster to be called. Clusters are then aligned between individuals, and only those found in a user specified minimum number of individuals are kept as loci. We tested clustering thresholds of 90-99% on a subset of the data, and chose 94% as the optimum based on the distribution of nucleotide diversity and total number of loci identified. We allowed a Phred quality score of <20 in a maximum of 4 sites per sequence. To avoid including singletons and spurious loci we kept clusters only if they had >5x coverage per individual, and were found in at least 4 individuals. Since the *de novo assembly* and variant calling step is the most time consuming, we applied minimal quality filters here to maximise the number of loci identified in the data. Further quality filtering is described in the next section.

SNP filtering and validation - Possible sequencing errors, uninformative SNPs, paralogs, and missing data were removed from the pyRAD variant file using the following filters: 1) SNPs with a minor allele count of 60 (that is 5.8% in our dataset) were removed, since alleles that occur at a low frequency are likely to be sequencing error, or to bias tests for selection (Roesti *et al.* 2012),

2) we retained only SNPs genotyped in more than 80% of individuals. Variants to remove were specified using the *--max-missing* function in VCFtools v0.1.12b (Danecek *et al.* 2011). 3) To reduce linkage in the dataset we retained only one SNP per 125 bp locus using the *--thin* function in VCFtools. 4) All individuals genotyped at less than 50% of the loci were removed using the *--remove* function in VCFtools. 5) We tested whether loci were in Hardy Weinberg Equilibrium (HWE) within each population using PLINK (Purcell *et al.* 2007), and removed any loci deviating from HWE or with an overall observed heterozygosity of more than 0.6.

1. Population structure

We visualized genetic distances among populations across Switzerland using a Principal Component Analysis (PCA) calculated using the *PCAdapt* v3.0.3 R package (Luu *et al.* 2016; R Core Team 2016). Next, we used two population clustering algorithms to determine the most likely number of genetic groups in the data. The first was a variational Bayesian framework implemented in fastStructure (Raj *et al.* 2014) to estimate the ancestry proportions for K = 2-5. The authors suggest that the most likely K (number of clusters) will fall within a range between the K that maximises the log-marginal likelihood and the smallest K that accounts for the majority of the ancestry in the sample.

The second clustering algorithm was a matrix factorisation method in TESS3 (Caye *et al.* 2016) to estimate the most likely number of clusters in the data while taking the geographic distribution of samples into account. TESS3 uses the law of total probability to estimate a Q-matrix of ancestry coefficients and G-matrix of genotype frequencies for a range of K. A least square minimisation algorithm for matrix factorisation is implemented, and introduces spatial constraints on the G-matrix to ensure that genotypes are more likely to be shared between individuals that are geographically closer together. The most likely K is based on cross-entropy scores calculated from results of the test dataset (here 5% of the data) with that of the training dataset.

Based on these results we defined four geographically distinct genetic clusters (Fig. 3.1). These regions differ in colonisation history and

mitochondrial genetic diversity (Chapter 1), so we expected them to differ in contemporary genetic structure as well. To investigate this, we conducted all further analyses independently for each region. The regions are named as follows: northern Switzerland (CH-North), southern Switzerland (CH-South) which is divided into an eastern (CH-SouthEast) and western (CH-SouthWest) region, and eastern Switzerland (CH-East). We use CH-South as an acronym for southern Switzerland (i.e. CH-SouthWest and CH-SouthEast together).

2. Is genetic diversity correlated with elevation?

Our second aim was to determine how genetic diversity is distributed across elevation. Specifically, we examined the relationship between genetic diversity and elevation and tested whether signals from mitochondrial and genomic data differ.

We calculated haplotype diversity (Hd) and nucleotide diversity (π) of the mitochondrial data for each population, using DnaSP 5.0 (Librado & Rozas 2009). For the SNP data, we calculated four measures of genetic diversity for each population: the frequency of heterozygous bases per individual across all sequenced bases (IUPAC codes), gene diversity (H_S), deviation from random mating (F_{IS}), and observed heterozygosity (H_O) (*hierfstat* v0.04-22, Goudet 2005; Jombart 2008; *adegenet* v2.0.1, Jombart & Ahmed 2011). Finally, we include the number of clutches found at each site as an estimate of current population size, which is appropriate because *R. temporaria* females typically oviposit a single egg mass per year.

We tested for geographic variation in the relationships between genetic diversity and elevation by regressing the five molecular measures of diversity (Hd, π, F_{IS}, H_O, H_S) and the number of clutches against elevation, geographic region (a categorical effect), and their interaction. Elevation was normalized before analysis (mean = 0, SD = 1). A second set of models that included the square of elevation and its interaction with region was also fitted to allow for non-linear relationships between diversity and elevation. We used the Akaike Information Criterion (AIC) to select the better of the two models for each response. Significance was evaluated using the *Anova* function in the *car* package (Fox & Weisberg 2011). To reduce the variance in genetic diversity

due to small sample size, populations with less than four individuals were excluded from these analyses.

For all the diversity measures that were significantly different between regions, we visualise their relationship with elevation by plotting the linear regression. In addition we regress the number of clutches found at each site against elevation as an estimate of the change in population size across elevation.

3. Isolation by distance and isolation by elevation

Our third aim was to determine whether genetic distance increases with geographic distance or elevation. Genetic distance based on SNP data was calculated as pairwise measures of Weir and Cockerham's F_{ST} (Weir & Cockerham 1984) between populations using the *pairwise.fst* function from the *hierfstat v.4-22* package in R (Goudet 2005).

We calculated the great circle geographic distance between all population pairs using the *rdist.earth* function in the *fields v.8.10* (Nychka *et al.* 2015) package in R. Elevation was estimated from topographic maps and from Google Earth v10.6.8.

We used a Mantel test to test for isolation by distance. Further, to account for the fact that geographic distance and elevation are confounded, we used partial mantel tests to test for isolation by distance and isolation by elevation. Mantel tests determine the linear correlation between two matrices. Partial Mantel tests remove the effects of a third matrix by determining the correlation between two matrices while conditioning the analysis on a third matrix (Legendre & Legendre 2012). The analyses were conducted on all the samples combined (CHall) and on each region separately. The analysis was implemented in the R package *vegan 2.4-2* (Oksanen *et al.* 2015).

4. Relative contribution of geographic distance and elevation to genetic divergence?

Our fourth aim was to determine how much genetic divergence between populations could be attributed to the geographic distance and the elevational difference between them. We compared results between mitochondrial and

SNP datasets to determine whether the effects have changed over time.

We used a form of multivariate regression, redundancy analysis (RDA), to estimate the relative contribution of elevation and geographic distance to genetic variation. Multivariate regressions perform better than linear regressions when predictor variables are correlated (Legendre & Fortin 2010; Legendre & Legendre 2012). As response variables we used mitochondrial haplotype diversity and the minor allele frequency across all the RAD loci. Minor allele frequencies were calculated using *vcftools* and Plink v.1.0.7. As explanatory variables we used the geographic and elevational distance from the geographically most distant low elevation population in each region. For CH-North: rade; CHS-SouthWest: orge; CHS-SouthEast: scai: CH-East: ente.

We used a series of RDA models to determine the total genetic variance, and the proportion explained by each of the predictor variables. The full RDA model included both elevation and geographic distance as explanatory variables. We then used partial RDA (pRDA) to partition the effects of one explanatory variable while estimating the effects of the second. In other words, pRDA1 estimates the effects of elevation while keeping the effects of geographic distance constant, and vice versa for pRDA2. The significance of each model was assessed with 999 permutations using the *anova* function in R. The joint contribution of elevation and geography was calculated as the residual variance of the full RDA model after removing the contributions of each partial RDA (i.e. RDAfull - (pRDA1 + pRDA2)). RDA was implemented with the *vegan v.2.4.2* package (Oksanen *et al.* 2015) in R.

Results

Genetic data

The SNP dataset comprised 992 individuals (each representing a family) from 82 populations (2-20 individuals per population) genotyped at 1827 loci with a genotyping rate of 89.6%. The mitochondrial dataset comprised 368 individuals (each representing a family) from 72 populations sequenced at 448bp of the *cytb* gene (Table S3.1)

1. Population structure

The principal components analysis (PCA) revealed that most genetic variation (46.4%) was captured by the first five components, with 29% explained by the first two axes. PC1 (19.5%) separated populations north and south of the alpine ridge bisecting Switzerland, and PC2 (9.3%) partitioned populations roughly from west to east (Fig. S3.1).

Clustering analyses suggested that between four (TESS3) and eight (fastStructure) clusters (K) best explained the data. Visualising the geographic distribution of these clusters (Fig. 3.1), we found that K=2 clustered populations north (CH-North) and south (CHS) of the alpine ridge. At K=3, eastern Switzerland (CH-East) clustered separately, while at K=4 southern Switzerland was separated into eastern (CHS-SouthEast) and western (CHS-SouthWest) groups. At K > 4 the cluster from eastern Switzerland was split up further, while CH-North, CHS-SouthEast, and CHS-SouthWest remained unchanged. Based on these results, we divided populations into four geographic regions. Individual assignment to regions was based on a combination of the fastStructure and TESS3 cluster assignments and geographic location of the population. We included all of eastern Switzerland in a single cluster even though the northernmost populations grouped genetically with CH-North and the southernmost populations grouped with CH-South. We did this for two reasons. Firstly, since eastern Switzerland represents a contact zone between the CH-North and CH-South genetic lineages, we wanted to include the full continuum of cluster assignment from

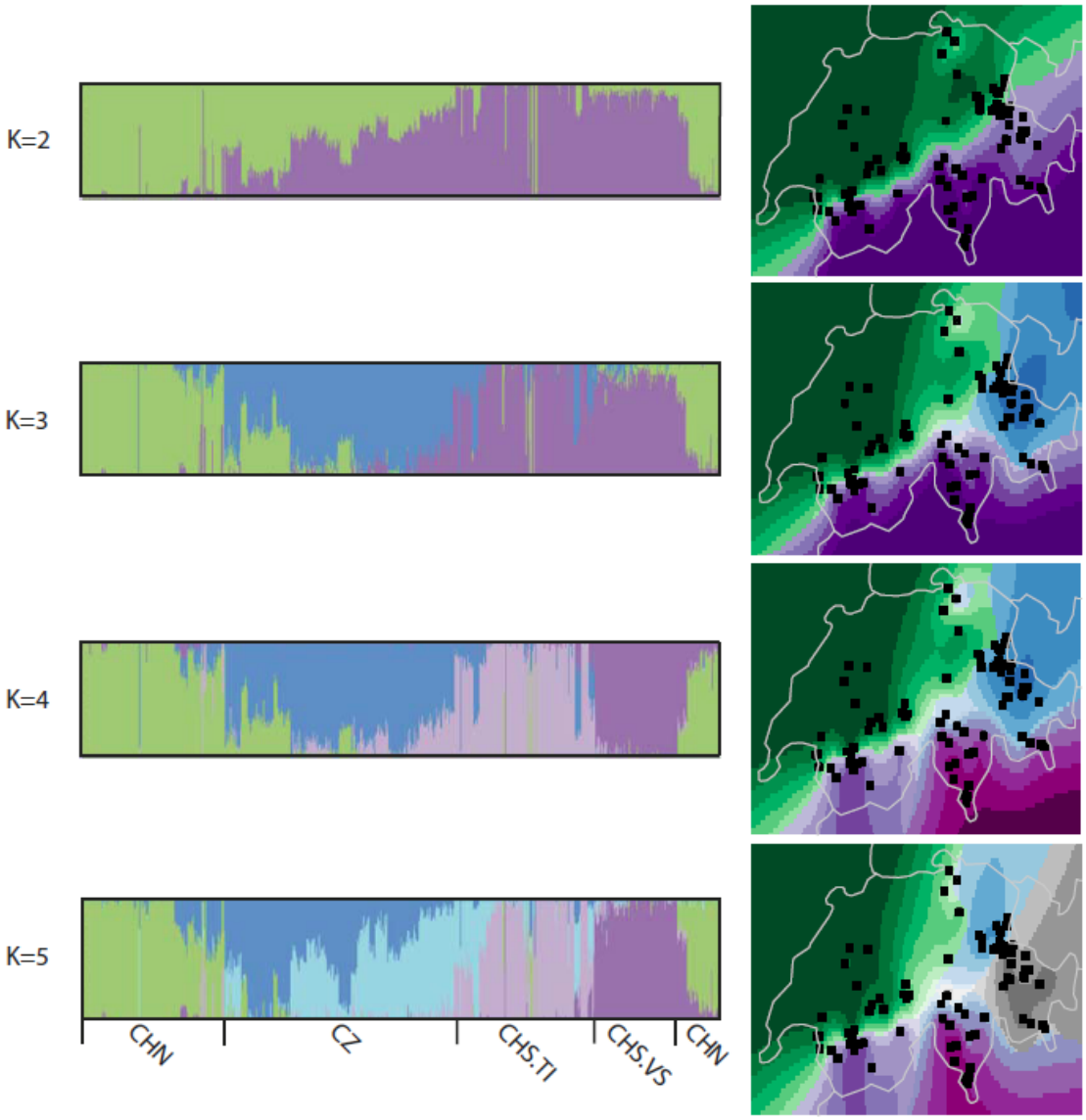


Figure 3.1 Population cluster analysis using fastStructure (left) and TESS3 (right). We show population assignment results assuming up to five genetic clusters (K=2-5). fastStructure uses an approximate Bayesian approach to estimate the posterior probability of each individuals assignment to the assumed number of clusters. The bar chart depicts each individual as a vertical line coloured according to the estimated population assignment. Individuals are ordered by geographic sampling region: CH-North (CHN) are ordered from west to east, CH-East (CZ) is ordered from north to south, CH-South are ordered from east to west, hence CH-SouthEast (CHS.TI) appears before CH-SouthWest (CHS.VS). Two populations from the CH-South geographic region assigned to the CH-North genetic group, hence the bar graph starts and ends with CH-N. TESS3 incorporates spatial weighting when determining assignment of individuals to genetic clusters. Clusters are coloured roughly by their geographic location and these colours are used in all following figures. At K=5 light blue in fastStructure = grey in TESS3.

CH-North to CH-South. Secondly, the southernmost populations from CH-East are not genetically the same as CHS-SoutEast. At K=5 these three populations group with southern CH-East, and at K=6 they form an independent cluster. Previous phylogeographic results suggest that this region of Switzerland was independently colonised from northern Italy (Chapter 1). One population located in central Switzerland (oalp) was removed from further analyses, as it is geographically isolated from all the clusters and had ambiguous assignments in the cluster analysis.

Pairwise population divergence (F_{ST}) was 0.14 on average, but ranged from 0.02 up to 0.37 (Fig. S3.2). The highest F_{ST} was between two high elevation populations from CH-North (stir) and CH-South (forn). Linear models indicated that most measures of genetic diversity (H_S and H_O; SNP data) and haplotype diversity (Hd and π; *cytb* data) varied significantly among regions (Fig. 3.2). Genetic diversity was lower on average in CH-North compared to CHS-SouthWest and CHS-SoutEast, and was generally higher in CH-East (Table S3.2). The mean frequency of heterozygous SNPs across all sequenced base-pairs did not differ among regions (Fig. S.3.3).

2. Is genetic diversity correlated with elevation?

Fig. 3.2 and Table 3.1 summarize relationships between genetic variation and elevation. In all cases, the linear relationship with elevation was found to fit the data better than the quadratic relationship. All diversity measures except F_{IS} were significantly different between regions. SNP genetic diversity (H_S) and mitochondrial haplotype diversity (Hd) were significant for region by elevation, suggesting that H_S and Hd are distributed differently across elevation in different regions.

To visualise the results we regressed each diversity measure (excluding F_{IS}) against elevation and determined the correlation coefficient in each case (Table 3.2). In CHS-SouthEast we found positive correlations between H_S ($p>0.05$) and elevation, and between H_O ($p<0.05$) and elevation. In the other regions H_S decreased with elevation (significantly in CHS-SouthWest only), while there was no or only a slight correlation between H_O and elevation. Based on the mitochondrial data, genetic diversity increased

significantly across elevation in CH-N (Hd: $R=0.58$, $p<0.05$; π: $R=0.57$, $p<0.05$). Genetic diversity remained unchanged or decreased across elevation in the other regions, but results were not significant. Clutch size decreased with elevation for all regions except CH-N, although all correlations were non-significant.

3. Isolation by distance and isolation by elevation

Mantel tests found significant isolation by distance in all datasets except CHS-SouthWest (Table 3.3). In testing for isolation by elevation, results from partial Mantel tests were non-significant for the full dataset, and all regions except CH-North (Table 3.3). These results suggest that gene flow decreases with geographic distance between populations, but that not with elevation.

4. Relative contribution of geographic distance and elevation to genetic divergence

Redundancy analysis (RDA) showed that 14-25 % of variation in allele frequency could be explained by geographic distance and elevation. The full RDA models explained the most variation at smaller geographic scales (CHS-SouthWest 25%; CHS-SouthEast 20%). Using partial RDA we found that geographic distance and elevation explained approximately the same amount of variation within each region. The combined effect of distance and elevation was evident only at larger geographic scales (Table S3a).

RDA explained 8-35% of mitochondrial haplotype diversity within each region. In contrast to the SNP data, elevation and the combination of geography and elevation explained most of the variance. The effect of geographic distance was negligible across all regions (Table S3b).

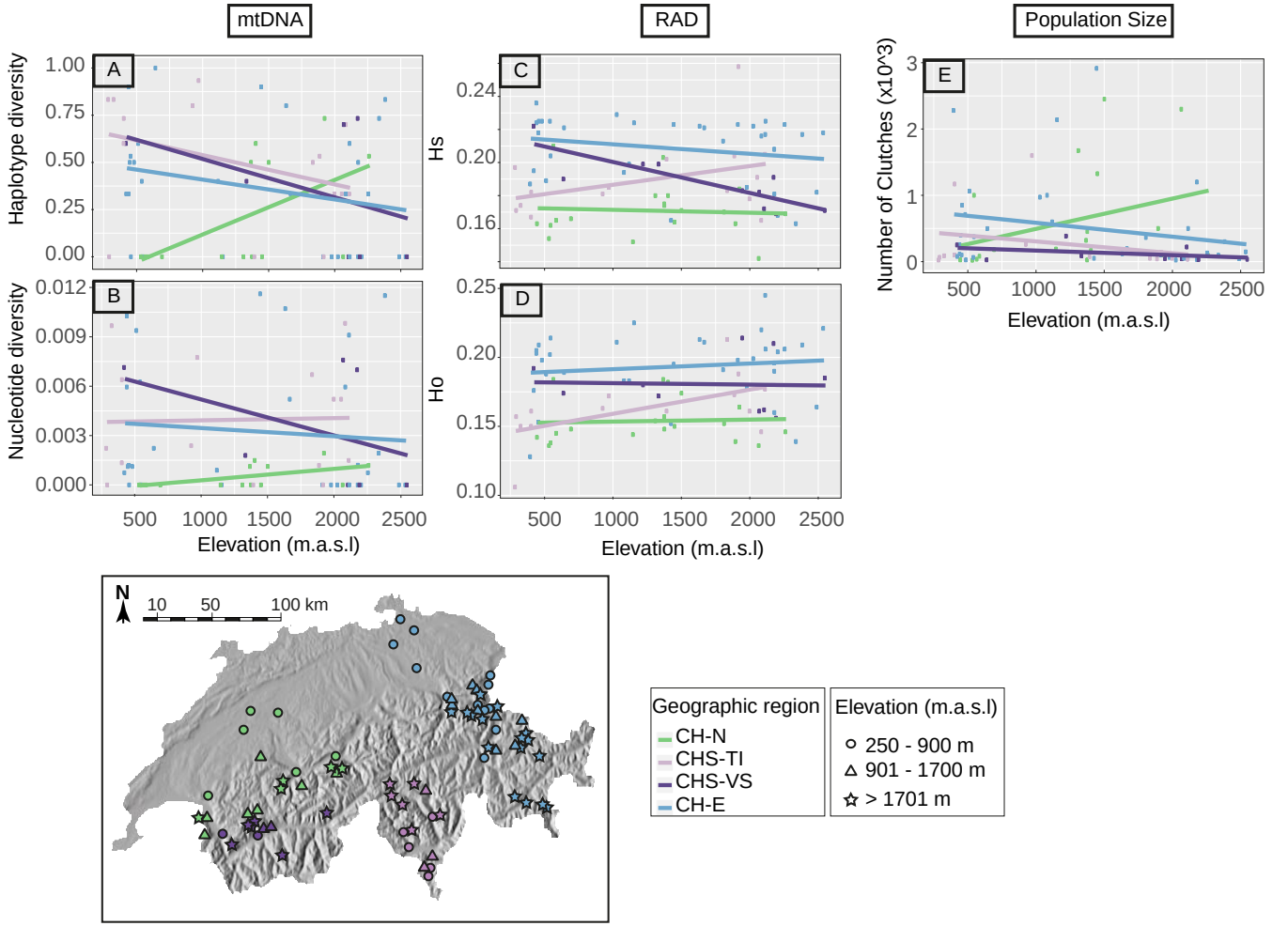


Figure 3.2 The change in genetic diversity across elevation calculated from mitochondrial *cytb* data (A-B), genome wide markers (C-D), and number of clutches (E). Populations and their elevations are indicated with shapes on the map, and coloured by geographic region (see Key and map). H_O: observed heterozygosity; H_S: genetic diversity.

Table 3.1 Results from anova of linear models used to assess the relationship between diversity measures and elevation. The first table shows results from SNP data (F_{IS} = inbreeding coefficient; H_S = genetic diversity; H_O = observed heterozygosity).The second table shows mitochondrial data (Hd = haplotype diversity; nd = nucleotide diversity). We tested for significant relationships with elevation, region, and their interaction as explanatory variables. All significant results are highlighted in grey. Significant results for region show that diversity measures differed between regions, and significant elevation x region show that the slope of the diversity measure across elevation was different between regions.

	SNP data								
Explanatory variable	**F_{IS}**			**H_S**			**Ho**		
	F value	**df**	**Significance**	**F value**	**df**	**Significance**	**F value**	**df**	**Significance**
elevation	9.56	1	**0.00**	1.07	1	0.31	2.81	1	0.10
region	1.50	3	0.22	17.73	3	**0.00**	**14.20**	**3**	**0.00**
elevation x region	0.44	3	0.73	2.89	3	**0.04**	1.05	3	0.38

	mtDNA					
Explanatory variable	**Hd**			**nd**		
	F value	**df**	**Significance**	**F value**	**df**	**Significance**
elevation	2.03	1	0.16	0.33	1	0.57
region	2.58	3	**0.06**	3.03	3	**0.04**
elevation x region	2.57	3	**0.06**	0.48	3	0.70

Table 3.2 Correlation coefficients and significance of five measures of genetic diversity regressed against elevation (see Fig. 2). Results are shown for each region and are based on mitochondrial (mtDNA) and genome-wide (SNP) markers. Significant results are highlighted in grey. Hd: haplotype diversity; π: nucleotide diversity; F_{IS}: fixation index; H_S: genetic diversity; H_O: observed heterozygosity.

	mtDNA				SNP					
Region	**Hd**		**π**		**H_S**		**H_O**		**Clutch size**	
	R^2	**p**	**R^2**	**p**	**R^2**	**p**	**R^2**	**p**	**R^2**	**p**
CH-North	0.58	**0.02**	0.57	**0.03**	-0.06	0.8	0.05	0.83	0.05	0.18
CHS-SouthEast	-0.37	0.17	0.03	0.91	0.38	0.17	0.57	**0.03**	0	0.34
CHS-SouthWest	-0.43	0.34	-0.43	0.34	-0.83	**0**	-0.04	0.91	0.03	0.3
CH-East	-0.26	0.17	-0.1	0.61	-0.22	0.2	0.13	0.47	0.03	0.18

Table 3.3 Results from Mantel and partial Mantel tests for isolation by distance and isolation by elevation. Analyses were conducted on the full dataset and on each region independently. Mantel correlations (r_M) and significance (P) are shown for all tests. Significant test results are shown in bold, and non-significant results are highlighted in grey.

Region	Matrix1	Matrix2	Control matrix	r_M	P
CH.all	Geography	Fst		0.30	**0.00**
	Elevation	Fst		0.05	**0.04**
	Geography	Fst	Elevation	0.30	**0.00**
	Elevation	Fst	Geography	0.04	0.06
CH-North	Geography	Fst		0.36	**0.00**
	Elevation	Fst		0.32	**0.00**
	Geography	Fst	Elevation	0.39	**0.00**
	Elevation	Fst	Geography	0.35	**0.00**
CH-East	Geography	Fst		0.64	**0.00**
	Elevation	Fst		0.15	**0.00**
	Geography	Fst	Elevation	0.63	**0.00**
	Elevation	Fst	Geography	-0.08	0.98
CH-SouthEast	Geography	Fst		0.23	**0.03**
	Elevation	Fst		0.14	**0.04**
	Geography	Fst	Elevation	0.19	**0.06**
	Elevation	Fst	Geography	0.06	0.25
CH-SouthWest	Geography	Fst		0.05	0.38
	Elevation	Fst		0.03	0.41
	Geography	Fst	Elevation	0.05	0.40
	Elevation	Fst	Geography	0.03	0.38

Discussion

Here we describe how the geographic distribution of genetic structure across the elevational distribution of a frog is influenced by both contemporary gene flow and colonisation history. Broad scale population structure mirrored the colonisation history of the Alps, and we identified four geographically separated genetic clusters. Overall we found a pattern of isolation by distance between populations. However, genetic divergence does not increase significantly across elevation, which suggests that the 2000-m elevational gradient sampled here was not a strong barrier to gene flow. Despite high levels of gene flow across elevation, we found that a relatively large proportion of the total genomic variation (4-13%) associated with elevation, indicative of adaptive divergence associated with the environmental gradient. We also find regional differences in the distribution of genetic diversity across elevation based on both historic (mtDNA) and contemporary (RAD) genetic markers. We discuss how regional differences in colonisation history and contemporary gene flow could result in these patterns, and what the implications might be for adaptation to elevational gradients.

Population structure mirrors colonisation history

Population structure of *R. temporaria* reflects the colonisation history of Switzerland and is governed by the geographic complexity of the Alps. Similar to the distribution of mitochondrial lineages (Chapter 1), we found two main genetic clusters separated by the alpine ridge bisecting Switzerland, with an extensive contact zone between them in eastern Switzerland. The alpine ridge bisecting Switzerland, particularly the section separating CHS-SouthWest and CH-North, culminates in an unbroken, steep and rocky ridge, thus confining populations to northern and southern Switzerland. However, gene flow is channeled along the low elevation valleys in both eastern and western Switzerland, as evidenced by the gradient in assignment to the CH-North and CH-South genetic population clusters in these regions. Similar patterns of colonisation and population structure have been reported for several plant species (Alvarez *et al.* 2009; Rogivue *et al.* 2018).

Populations from CHS-SouthWest and CHS-SouthEast are also separated by a high mountain ridge, and were assigned to different genetic clusters based on genome-wide markers. Previous work (Chapter 1) suggested, but did not demonstrate, that these two regions were colonized separately from different microrefugia in the northern Italian Alps. However, our SNP data give little evidence of contemporary gene flow between CH-SouthEast and CH-SouthWest, indicating that current geographic isolation may account for different cluster assignments. There were almost no individuals that assigned to both CHS-SouthEast and CHS-SouthWest, whereas both regions showed introgression with CH-North, providing evidence of gene flow between these lineages in north eastern CHS-SouthEast and south western CHS-SouthWest.

Isolation by distance but not by elevation

We found that genetic distance increases significantly with geographic distance, but that elevation does not pose a significant barrier to dispersal for *R. temporaria* in the Swiss Alps. High levels of gene flow (F_{ST} 0.02) have previously been reported across an elevational gradient of about ~700 m in Scotland (Muir *et al.* 2014b), and at a small geographic scales in the French Alps (Bonin *et al.* 2006). Features such as elevational gradient or slope and topographic relief have been found to be important determinants of gene-flow across elevation in several amphibian species (e.g. Dudaniec *et al.* 2012; Wang 2012; Garcia *et al.* 2017). Thus, our result is surprising given the sheer elevational difference and ruggedness of the terrain in the Alps. It also implies that there is little habitat selection or selection against immigrants across elevation, despite the dramatic difference between high and low elevation environments.

The redundancy analysis (RDA) – showing that elevation explains as much genomic variation as isolation by distance does – strongly suggests some degree of adaptive divergence across the elevational gradient. Elsewhere, *R. temporaria* populations show adaptive divergence across elevation, with high- and low-elevation populations differing in larval performance traits (Miaud & Merilä 2001; Muir *et al.* 2013a). Similar adaptive

divergence has been found in the Swiss Alps (Bachmann 2017). Together these results show that adaptive divergence is maintained in the face of high gene flow across elevation in the Swiss Alps.

Genetic diversity decreases across elevation, with some exceptions

Genetic diversity is predicted to decrease with distance from glacial refugia due to repeated bottlenecks experienced by populations colonising new habitats (Hewitt 2000). Since colonisation must have proceeded from low to high elevation as glaciers retreated, we similarly expect a decrease in mitochondrial genetic diversity across elevation. This pattern was found in CHS-SouthWest, CH-SouthEast, and CH-East (Fig. 2A), and elevation explained the majority of the explainable mitochondrial variation within all four regions (Table S2). However, mitochondrial haplotype and nucleotide diversity increased with elevation in CH-North, while genomic diversity increased with elevation in CHS-SouthEast. These patterns could be due to cryptic refugia (nunataks) at high elevations in the Alps or population expansion. Nunataks are sheer rocky outcrops that remained free of snow and ice during the last glacial maximum. They are generally known to have harboured some Alpine plant lineages (e.g. Stehlik *et al.* 2002), but they would not have provided a hospitable habitat to an amphibian. We suggest that the pattern in CH-North reflects a population expansion during the colonisation of the elevational gradient. CH-North is thought to have colonised northern Switzerland from a refugium in northern Italy in a colonisation wave that moved around the west of Switzerland, through France, and into Switzerland from the north or north west (Chapter 1). Much of the genetic diversity was lost during the colonisation process, and mitochondrial haplotype and nucleotide diversities are much lower in CH-North than in CH-South. Moreover, the haplotype network in northern Switzerland is star-like, which is characteristic of new haplotypes that arise during population expansion (Chapter 1). The relatively high mitochondrial genetic diversity found in CHS-SouthWest and CHS-SouthEast, and their proximity to the glacial refugia suggest that population sizes were larger and colonisation of the elevational gradient proceeded without a population expansion.

The increased SNP genetic diversity at high elevations in CHS-SouthEast could be due to the southward colonisation of CH-North populations into CHS-SouthEast across the central alpine passes. We find evidence for this in the distribution of the mitochondrial haplotypes (Chapter 1), the admixed population assignments to CH-East and CHS-SouthEast clusters based on genomic data (Fig. 1), and the increase in heterozygosity across elevation. High gene flow across elevation in the Alps has been reported for a sessile oak (Alberto *et al.* 2010) and vole species (Cornetti *et al.* 2016). Moreover, similar trans-alpine colonisation routes and contact zones between lineages has been reported in several Alpine plant species (e.g. Rogivue *et al.* 2018)

Conclusions

We conclude that colonisation history, geographic distance, and adaptive divergence across elevation together determine the distribution of genomic variation across *R. temporaria* populations in Switzerland. Broadly, neutral population structure is determined largely by the colonisation and the topographic complexity of Switzerland that determines contemporary levels of gene flow between regions. In contrast, we found high levels of gene flow across elevation despite presenting a steep geographic and environmental gradient. At smaller geographic scales differences in demographic histories and contemporary gene flow have resulted in regional differences in the distribution of genetic diversity across elevation.

References

Akaike H (1973) Information theory and an extension of the maximum likelihood principle. In: *2nd International Symposium on Information Theory, Tsahkadsor, Armenia, USSR*, pp. 267–281. Budapest:Akadémiai Kiadó.

Alberto F, Niort J, Derory J *et al.* (2010) Population differentiation of sessile oak at the altitudinal front of migration in the French Pyrenees. *Molecular Ecology*, **19**, 2626–2639.

Alton LA, Franklin CE (2017) Drivers of amphibian declines: effects of ultraviolet radiation and interactions with other environmental factors. *Climate Change Responses*, **4**, DOI 10.1186/s40665-017-0034-7.

Altschup SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic Local Alignment Search Tool. *Journal of Molecular Biology*, **205**, 403–410.

Alvarez N, Thiel-Egenter C, Tribsch A *et al.* (2009) History or ecology? Substrate type as a major driver of spatial genetic structure in Alpine plants. *Ecology Letters*, **12**, 632–640.

Ammann B, Birks HJB, Brooks SJ *et al.* (2000) Quantification of biotic responses to rapid climatic changes around the Younger Dryas — a synthesis. *Paleogeography, Paleoclimatology, Paleoecology*, **159**, 313–347.

Avise JC, Riddle B (2009) Phylogeography : Retrospect and Prospect. *Journal of Biogeography*, **36**, 3–15.

Bachmann JC (2017) Adaptive Divergence across an Elevational Gradient in the Common Frog (Rana temporaria). University of Zurich.

Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, **30**, 2114–2120.

Bonin A (2008) Population genomics: a new generation of genome scans to bridge the gap with functional genomics. *Molecular ecology*, **17**, 3583–4.

Bonin A, Taberlet P, Miaud C, Pompanon F (2006) Explorative genome scan to detect candidate loci for adaptation along a gradient of altitude in the common frog (Rana temporaria). *Molecular biology and evolution*, **23**, 773–83.

Borcard D, Legendre P (2002) All-scale spatial analysis of ecological data by means of principal coordinates of neighbour matrices. *Ecological Modelling*, **153**, 51–68.

Braaker S, Heckel G (2009) Transalpine colonisation and partial phylogeographic erosion by dispersal in the common vole (Microtus arvalis). *Molecular Ecology*, **18**, 2528–2531.

Brady LD, Griffiths RA (2000) Developmental responses to pond desiccation in tadpoles of the British anuran amphibians (Bufo bufo, B. calamita and Rana temporaria). *Journal of Zoology, London*, **252**, 61–69.

Van Buskirk J (2012) Permeability of the landscape matrix between amphibian breeding sites. *Ecology and evolution*, **2**, 3160–7.

Canestrelli D, Cimmaruta R, Costantini V, Nascetti G (2006) Genetic diversity and phylogeography of the Apennine yellow-bellied toad Bombina pachypus, with implications for conservation. *Molecular Ecology*, **15**, 3741–3754.

Canestrelli D, Cimmaruta R, Nascetti G (2008) Population genetic structure and diversity of the Apennine endemic stream frog, Rana italica – insights on the Pleistocene evolutionary history of the Italian peninsular biota. *Molecular Ecology*, **17**, 3856–3872.

Canestrelli D, Nascetti G (2008) Phylogeography of the pool frog Rana (Pelophylax) lessonae in the Italian peninsula and Sicily: Multiple refugia, glacial expansions and nuclear-mitochondrial discordance. *Journal of Biogeography*, **35**, 1923–1936.

Cano JM, Laurila A, Palo J, Merilä J (2004) Population differentiation in G matrix structure due to natural selection in Rana temporaria. *Evolution*, **58**, 2013–2020.

Catchen JM, Amores A, Hohenlohe P, Cresko W, Postlethwait JH (2011) Stacks: Building and Genotyping Loci De Novo From Short-Read Sequences. *G3:Genes, Genomes, Genetics*, **1**, 171–182.

Caye K, Deist TM, Martins H, Michel O, François O (2016) TESS3: Fast inference of spatial population structure and genome scans for selection. *Molecular Ecology Resources*, **16**, 540–548.

CH2014–Impacts (2014) *Toward quantitative scenarios of climate change Impacts in Switzerland*. OCCR, FOEN, MeteoSwiss, C2SM, Agroscope, and ProClim, Bern, Switzerland.

Clement M, Posada D, Crandall KA (2000) TCS : a computer program to estimate gene genealogies. *Molecular Ecology*, **9**, 1657–1659.

Cornetti L, Lemoine M, Hilfiker D *et al.* (2016) Higher genetic diversity on mountain tops: the role of historical and contemporary processes in shaping genetic variation in the bank vole. *Biological Journal of the Linnean Society*, **118**, 233–244.

Croteau MC, Davidson MA, Lean DRS, Trudeau VL (2008) Global Increases in Ultraviolet B Radiation : Potential Impacts on Amphibian Development and Metamorphosis. *Physiological and Biochemical Zoology*, **81**, 743–761.

Crottini A, Andreone F, Kosuch J *et al.* (2007) Fossorial but widespread: the phylogeography of the common spadefoot toad (Pelobates fuscus), and the role of the Po Valley as a major source of genetic variability. *Molecular Ecology*, **16**, 2734–2754.

Danecek P, Auton A, Abecasis G *et al.* (2011) The variant call format and VCFtools. *Bioinformatics*, **27**, 2156–2158.

Dansgaard W, Johnsen SJ, Clausen HB *et al.* (1993) Evidence for general instability of past climate from a 250-kyr ice-core record. *Nature*, **364**, 218–220.

Darnault R, Rolland Y, Braucher R *et al.* (2011) Timing of the last deglaciation revealed by receding glaciers at the Alpine-scale : impact on mountain geomorphology. *Quaternary Science Reviews*, 1–16.

Darriba D, Taboada GL, Doallo R, Posada D (2012) jModelTest 2: more models, new heuristics and parallel computing. *Nature Methods*, **9**, 772.

Debieu M, Tang C, Stich B *et al.* (2013) Co-Variation between seed dormancy, growth rate and flowering time changes with latitude in Arabidopsis thaliana. *PloS one*, **8**, 1–12.

Demesure B, Comps B, Petit RJ (1996) Chloroplast DNA Phylogeography of the Common Beech (Fagus sylvatica L.) in Europe. *Evolution*, **50**, 2515–2520.

Drummond AJ, Suchard MA, Xie D, Rambaut A (2012) Bayesian Phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution*, **29**, 1969–1973.
Dudaniec RY, Spear SF, Richardson JS, Storfer A (2012) Current and historical drivers of landscape genetic structure differ in core and peripheral salamander populations. *PloS one*, **7**.
Eaton DAR (2014) PyRAD: Assembly of de novo RADseq loci for phylogenetic analyses. *Bioinformatics*, **30**, 1844–1849.
Ellis N, Smith SJ, Pitcher CR (2012) Gradient forests : calculating importance gradients on physical predictors. *Ecology*, **93**, 156–168.
Excoffier L, Lischer HEL (2010) An Integrated Software Package for Population Genetics Data Analysis. *Molecular Ecology Resources*, **10**, 564–567.
Fitzpatrick SW, Gerberich JC, Kronenberger JA, Angeloni LM, Funk WC (2015) Locally adapted traits maintained in the face of high gene flow. *Ecology Letters*, **18**, 37–47.
Fitzpatrick MC, Keller SR (2015) Ecological genomics meets community-level modelling of biodiversity: mapping the genomic landscape of current and future environmental adaptation. *Ecology Letters*, **18**, 1–16.
Forester BR, Jones MR, Joost S, Landguth EL, Lasky JR (2016) Detecting spatial genetic signatures of local adaptation in heterogeneous landscapes. *Molecular Ecology*, **25**, 104–120.
Fox J, Weisberg S (2011) *An {R} Companion to Applied Regression*. Thousand Oaks CA: Sage.
Freedman AH, Thomassen HA, Buermann W, Smith TB (2010) Genomic signals of diversification along ecological gradients in a tropical lizard. *Molecular ecology*, **19**, 3773–3788.
Frichot E, François O (2015) LEA : An R package for landscape and ecological association studies (B O'Meara, Ed,). *Methods in Ecology and Evolution*, **6**, 925–929.
Frichot E, Schoville SD, Bouchard G, François O (2013) Testing for associations between loci and environmental gradients using latent factor mixed models. *Molecular Biology and Evolution*, **30**, 1687–99.
Frichot E, Schoville S, de Villemereuil P, Gaggiotti OE, François O (2015) Detecting adaptive evolution based on association with ecological gradients: Orientation matters! *Heredity*, **115**, 22–28.
Garcia VOS, Ivy C, Fu J (2017) Syntopic frogs reveal different patterns of interaction with the landscape: A comparative landscape genetic study of Pelophylax nigromaculatus and Fejervarya limnocharis from central China. *Ecology and Evolution*, **7**, 9294–9306.
Gosner KL (1960) A Simplified Table for Staging Anuran Embryos and Larvae with Notes on Identification. *Herpetologica*, **16**, 183–190.
Goudet J (2005) Hierfstat, a package for R to compute and test hierarchical F-statistics. *Molecular Ecology Notes*, **5**, 184–186.
Gugerli F, Englisch T, Niklfeld H *et al.* (2008) Relationships among levels of biodiversity and the relevance of intraspecific diversity in conservation – a project synopsis. *Perspectives in Plant Ecology, Evolution and Systematics*, **10**, 259–281.
Günther T, Coop G (2013) Robust identification of local adaptation from allele frequencies. *Genetics*, **195**, 205–220.

Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series*, **41**, 95–98.

Harrisson KA, Amish SJ, Pavlova A *et al.* (2017) Signatures of polygenic adaptation associated with climate across the range of a threatened fish species with high genetic connectivity. *Molecular ecology*, **26**, 6253–6269.

Hecht BC, Matala AP, Hess JE, Narum SR (2015) Environmental adaptation in Chinook salmon (Oncorhynchus tshawytscha) throughout their North American range. *Molecular Ecology*, **24**, 5573–5595.

Heiri C, Bugmann H, Tinner W, Heiri O, Lischke H (2006) A model-based reconstruction of Holocene treeline dynamics in the Central Swiss Alps. *Journal of Ecology*, **94**, 206–216.

Hermisson J, Pennings PS (2005) Soft sweeps: molecular population genetics of adaptation from standing genetic variation. *Genetics*, **169**, 2335–52.

Hewitt GM (1996) Some genetic consequences of ice ages, and their role in divergence and speciation. *Biological Journal of the Linnean Society*, **58**, 247–276.

Hewitt G (1999) Post-glacial re-colonization of European biota. *Biological Journal of the Linnean Society*, **68**, 87–112.

Hewitt G (2000) The genetic legacy of the Quaternary ice ages. *Nature*, **405**, 907–913.

Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A (2005) Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, **25**, 1965–1978.

Hitchings SP, Beebee TJC (1997) Genetic substructuring as a result of barriers to gene flow in urban Rana temporaria (common frog) populations: implications for biodiversity conservation. *Heredity*, **79**, 117–127.

Hjernquist MB, Soderman F, Jonsson KI *et al.* (2012) Seasonality determines patterns of growth and age structure over a geographic gradient in an ectothermic vertebrate. *Oecologia*, **170**, 641–649.

Hoban S, Kelley JL, Lotterhos KE *et al.* (2016) Finding the Genomic Basis of Local Adaptation: Pitfalls, Practical Solutions, and Future Directions. *The American Naturalist*, **188**, 000–000.

Hoffmann A a, Sgrò CM (2011) Climate change and evolutionary adaptation. *Nature*, **470**, 479–85.

Huelsenbeck JP, Ronquist F (2001) MRBAYES : Bayesian inference of phylogenetic trees. *Bioinformatics Application Note*, **17**, 754–755.

Ilut DC, Nydam ML, Hare MP (2014) Defining loci in restriction-based reduced representation genomic data from nonmodel species: Sources of bias and diagnostics for optimal clustering. *BioMed Research International*, **2014**, 9 pages.

Jombart T (2008) Adegenet: A R package for the multivariate analysis of genetic markers. *Bioinformatics*, **24**, 1403–1405.

Jombart T, Ahmed I (2011) adegenet 1.3-1: New tools for the analysis of genome-wide SNP data. *Bioinformatics*, **27**, 3070–3071.

Jones MR, Forester BR, Teufel AI *et al.* (2013) Integrating landscape genomics and spatially explicit approaches to detect loci under selection in clinal populations. *Evolution*, **67**, 3455–68.

Kawecki TJ, Ebert D (2004) Conceptual issues in local adaptation. *Ecology Letters*, **7**, 1225–1241.
Lasky JR, Des Marais DL, McKay J *et al.* (2012) Characterizing genomic variation of Arabidopsis thaliana: the roles of geography and climate. *Molecular Ecology*, **21**, 5512–5529.
Laugen AT, Kruuk LEB, Laurila A *et al.* (2005a) Quantitative genetics of larval life-history traits in Rana temporaria in different environmental conditions. *Genetics Research*, **86**, 161–170.
Laugen AT, Laurila A, Jönsson KI, Söderman F, Merilä J (2005b) Do common frogs (*Rana temporaria*) follow Bergmann's rule? *Evolutionary Ecology Research*, **7**, 717–731.
Laugen AT, Laurila A, Merilä J (2002) Maternal and genetic contributions to geographical variation in Rana temporaria larval life-history traits. *Biological Journal fo the Linnean Society*, **76**, 61–70.
Laugen AT, Laurila A, Merilä J (2003a) Latitudinal and temperature-dependent variation in embryonic development and growth in Rana temporaria. *Oecologia*, **135**, 548–554.
Laugen AT, Laurila A, Rasanen K, Merilä J (2003b) Latitudinal countergradient variation in the common frog (Rana temporaria) development rates – evidence for local adaptation. *Journal of Evolutionary Biology*, **16**, 996–1005.
Laurila A, Karttunen S, Meril J (2002) Adaptive Phenotypic Plasticity and Genetics of Larval Life Histories in Two Rana Temporaria Populations. *Evolution*, **56**, 617–627.
Laurila A, Pakkasmaa S, Merila J (2001) Influence of Seasonal Time Constraints on Growth and Development of Common Frog Tadpoles: A Photoperiod Experiment. *Oikos*, **95**, 451–460.
Legendre P, Fortin M-J (2010) Comparison of the Mantel test and alternative approaches for detecting complex multivariate relationships in the spatial analysis of genetic data. *Molecular Ecology Resources*, **10**, 831–844.
Legendre P, Legendre LF (2012) *Numerical ecology*. Elsevier.
Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics Application Note*, **25**, 1451–1452.
Lindgren B, Laurila A (2005) Proximate causes of adaptive growth rates: growth efficiency variation among latitudinal populations of Rana temporaria. *Journal of Evolutionary Biology*, **18**, 820–828.
Lindgren B, Laurila A (2009) Physiological variation along a geographical gradient: is growth rate correlated with routine metabolic rate in Rana temporaria tadpoles? *Biological Journal of the Linnean Society*, **98**, 217–224.
Loman J, Claesson D (2003) Plastic response to pond drying in tadpoles Rana temporaria: tests of cost models. *Evolutionary Ecology Research*, **5**, 179–194.
Lugon-Moulin N, Hausser J (2002) Phylogeographical structure, postglacial recolonization and barriers to gene flow in the distinctive Valais chromosome race of the common shrew (Sorex araneus). *Molecular Ecology*, **11**, 785–794.
Luikart G, England PR, Tallmon D, Jordan S, Taberlet P (2003) The power and promise of population genomics: from genotyping to genome typing.

Nature Reviews Genetics, **4**, 981–94.
Luu K, Bazin E, Blum MGB (2016) pcadapt: An R package to perform genome scans for selection based on principal component analysis. *Molecular Ecology Resources*, **33**, 67–77.
Manel S, Perrier C, Pratlong M *et al.* (2016) Genomic resources and their influence on the detection of the signal of positive selection in genome scans. *Molecular ecology*, **25**, 170–184.
Manel S, Poncet BN, Legendre P, Gugerli F, Holderegger R (2010) Common factors drive adaptive genetic variation at different spatial scales in Arabis alpina. *Molecular ecology*, **19**, 3824–35.
Marquis O, Miaud C (2008) Variation in UV sensitivity among common frog Rana temporaria populations along an altitudinal gradient. *Zoology*, **111**, 309–317.
Marquis O, Miaud C, Lena J-P (2008) Developmental responses to UV-B radiation in common frog Rana temporaria embryos from along an altitudinal gradient. *Population Ecology*, **50**, 123–130.
Mátyás G, Sperisen C (2001) Chloroplast DNA polymorphisms provide evidence for postglacial re-colonisation of oaks (Quercus spp.) across the Swiss Alps. *Theoretical and Applied Genetics*, **102**, 12–20.
McCain CM, Colwell RK (2011) Assessing the threat to montane biodiversity from discordant shifts in temperature and precipitation in a changing climate. *Ecology Letters*, **14**, 1236–1245.
Merila J, Laurila A, Laugen AT, Rasanen K, Pahkala M (2000) Plasticity in age and size at metamorphosis in Rana temporaria: comparison of high and low latitude populations. *Ecography*, **23**, 457–465.
Merilä J, Laurila A, Laugen AT, Räsänen K, Pahkala M (2000) Plasticity in Age and Size at Metamorphosis in Rana Temporaria: Comparison of High and Low Latitude Populations. *Ecography*, **23**, 457–465.
Messer PW, Petrov DA (2013) Population genomics of rapid adaptation by soft selective sweeps. *Trends in Ecology and Evolution*, **28**, 659–669.
Miaud C, Guyetant R, Elmberg J (1999) Variations in life-history traits in the common frog Rana temporaria (Amphibia: Anura): a literature review and new data from the French Alps. *Journal of Zoology*, **249**, 61–73.
Miaud C, Merilä J (2001) Local adaptation or environmental induction? Causes of population differentiation in alpine amphibians. *Biota*, **2**, 31–50.
Muir AP, Biek R, Mable BK (2014a) Behavioural and physiological adaptations to low-temperature environments in the common frog, Rana temporaria. *BMC Evolutionary Biology*, **14**, 1471–2148.
Muir A, Piek R, Thomas R, Mable B (2014b) Local adaptation with high gene flow: temperature parameters drive adaptation to altitude in the common frog (Rana temporaria). *Molecular ecology*, **23**, 561–574.
Nychka D, Furrer R, Paige J, Sain S (2015) fields: Tools for spatial data.
Oksanen J, Blanchet FG, Kindt R *et al.* (2015) Vegan: community ecology package. R package vegan, vers. 2.2-1.
Pahkala M, Laurila A, Meril J (2002) Effects of ultraviolet-B radiation on common frog Rana temporaria embryos from along a latitudinal gradient. *Oecologia*, **133**, 458–465.
Pahkala M, Laurila A, Merila J (2000) Ambient Ultraviolet-B radiation reduces hatchling size in the common frog Rana temporaria. *Ecography*, **23**, 531–

538.
Pahkala M, Laurila A, Merila J (2001) Carry-over effects of ultraviolet-B radiation on larval tness in Rana temporaria. *Proceedings of the Royal Society of London B: Biological Sciences*, **268**, 1699–1706.
Pahkala M, Merila J, Ots I, Laurila A (2003) Effects of ultraviolet-B radiation on metamorphic traits in the common frog Rana temporaria. *Journal of the Zoological Society of London*, **259**, 57–62.
Palo JU, O'Hara RB, Laugen AT *et al.* (2003a) Latitudinal divergence of common frog (Rana temporaria) life history traits by natural selection: evidence from a comparison of molecular and quantitative genetic data. *Molecular ecology*, **12**, 1963–1978.
Palo JU, O'Hara RB, Laugen AT *et al.* (2003b) Latitudinal divergence of the common frog (Rana temporaria) life history traits by natural selection: evidence from a comparison of molecular quantitative genetic data. *Molecular Ecology*, **12**, 1963–1978.
Palo JU, Schmeller DS, Laurila A *et al.* (2004) High degree of population subdivision in a widespread amphibian. *Molecular Ecology*, **13**, 2631–2644.
Parisod C (2008) Postglacial recolonisation of plants in the western Alps of Switzerland. *Botanica Helvetica*, **118**, 1–12.
Parmesan C (2006) Ecological and Evolutionary Responses to Recent Climate Change. *Annual Review of Ecology, Evolution, and Systematics*, **37**, 637–669.
Parmesan C, Yohe G (2003) A globally coherent fingerprint of climate change impacts across natural systems. *Nature*, **421**, 37–42.
Pennings PS, Hermisson J (2006) Soft sweeps II - Molecular population genetics of adaptation from recurrent mutation or migration. *Molecular Biology and Evolution*, **23**, 1076–1084.
Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE (2012) Double Digest RADseq: An Inexpensive Method for De Novo SNP Discovery and Genotyping in Model and Non-Model Species. *PloS one*, **7**.
Polechová J, Barton NH (2015) Limits to adaptation along environmental gradients. *Proceedings of the National Academy of Sciences of the United States of America*, **112**, 6401–6406.
Poncet BN, Herrmann D, Gugerli F *et al.* (2010) Tracking genes of ecological relevance using a genome scan in two independent regional population samples of Arabis alpina. *Molecular ecology*, **19**, 2896–907.
Pritchard JK, Di Rienzo A (2010) Adaptation - not by sweeps alone. *Nature Reviews Genetics*, **11**, 665–7.
Purcell S, Neale B, Todd-Brown K *et al.* (2007) PLINK: A tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics*, **81**, 559–575.
Raj A, Stephens M, Pritchard JK (2014) FastSTRUCTURE: Variational inference of population structure in large SNP data sets. *Genetics*, **197**, 573–589.
Rehm EM, Olivas P, Stroud J, Feeley KJ (2015) Losing your edge: climate change and the conservation value of range-edge populations. *Ecology and evolution*, **5**, 4315–4326.
Rellstab C, Fischer MC, Zoller S *et al.* (2016) Local adaptation (mostly) remains local: reassessing environmental associations of climate-related

candidate SNPs in Arabidopsis halleri. *Heredity*, 1–9.
Rellstab C, Gugerli F, Eckert AJ, Hancock AM, Holderegger R (2015) A practical guide to environmental association analysis in landscape genomics. *Molecular Ecology*, **24**, 4348–70.
Rigling A, Bigler C, Eilmann B *et al.* (2013) Driving factors of a vegetation shift from Scots pine to pubescent oak in dry Alpine forests. *Global Change Biology*, **19**, 229–240.
Rodrigues N, Betto-Colliard C, Jourdan-Pineau H, Perrin N (2013) Within-population polymorphism of sex-determination systems in the common frog (Rana temporaria). *Journal of evolutionary biology*, **26**, 1569–1577.
Roesti M, Salzburger W, Berner D (2012) Uninformative polymorphisms bias genome scans for signatures of selection. *BMC Evolutionary Biology*, **12**, 94.
Roff DA (1996) The evolution of threshold traits in animals. *The Quarterly Review of Biology*, **71**, 3–35.
Rogers AR, Harpending H (1992) Population Growth Makes Waves in the Distribution of Pairwise Genetic Differences. *Molecular biology and evolution*, **9**, 552–569.
Rogivue A, Graf R, Parisod C, Holderegger R, Gugerli F (2018) The phylogeographic structure of Arabis alpina in the Alps shows consistent patterns across different types of molecular markers and geographic scales. *Alpine Botany*, **0**, 0.
Rousset F (1997) Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. *Genetics*, **145**, 1219–1228.
Roy K, Valentine JW, Jablonski D, Kidwell SM (1996) Scales of climatic variability and time averaging in Pleistocene biotas: implications for ecology and evolution. *Trends in Ecology & Evolution*, **11**, 458–463.
Schmeller DS, Palo JU, Merilä J (2008) A contact zone between two distinct Rana temporaria lineages in northern Germany. *Alytes*, **25**, 93–98.
Schneider S, Excoffier L (1999) Estimation of Past Demographic Parameters From the Distribution of Pairwise Differences When the Mutation Rates Vary Among Sites: Application to Human Mitochondrial DNA. *Genetics*, **152**, 1079–1089.
Schweizer RM, VonHoldt BM, Harrigan R *et al.* (2016) Genetic subdivision and candidate genes under selection in North American grey wolves. *Molecular Ecology*, **25**, 380–402.
Sillero N, Campos J, Bonardi A *et al.* (2014) Updated distribution and biogeography of amphibians and reptiles of Europe. *Amphibia-Reptilia*, **35**, 1–31.
Sork VL, Aitken SN, Dyer RJ *et al.* (2013) Putting the landscape into the genomics of trees: approaches for understanding local adaptation and population responses to changing climate. *Tree Genetics & Genomes*, **9**, 901–911.
Ståhlberg F, Olsson M, Uller T (2001) Population divergence of developmental thermal optima in Swedish common frogs, Rana temporaria. *Journal of Evolutionary Biology*, **14**, 755–762.
Stamatakis A (2006) RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics Application Note*, **22**, 2688–2690.
Stapley J, Reger J, Feulner PGD *et al.* (2010) Adaptation genomics: the next

generation. *Trends in ecology & evolution*, **25**, 705–12.
Stefani F, Gentilli A, Sacchi R *et al.* (2012) Refugia within refugia as a key to disentangle the genetic pattern of a highly variable species: The case of Rana temporaria Linnaeus, 1758 (Anura, Ranidae). *Molecular Phylogenetics and Evolution*, **65**, 718–726.
Stehlik I, Blattner FR, Holderegger R, Bachmann K (2002) Nunatak survival of the high Alpine plant Eritrichium nanum (L.) Gaudin in the central Alps during the ice ages. *Molecular Ecology*, **11**, 2027–2036.
Taberlet P, Fumagalli L, Wust-Saucy A-G, Cossons J-F (1998) Comparative phylogeography and postglacial colonization routes in Europe. *Molecular Ecolgy*, **7**, 453–464.
Teacher AGF, Garner TWJ, Nichols RA (2009) European phylogeography of the common frog (Rana temporaria): routes of postglacial colonization into the British Isles, and evidence for an Irish glacial refugium. *Heredity*, **102**, 490–6.
Templeton AR, Crandall KA, Sing CF (1992) Cladistic Analysis of Phenotypic Associations With Haplotypes Inferred From Restriction Endonuclease Mapping and DNA Sequence Data. III. Cladogram Estimation. *Genetics*, **132**, 619–633.
Thomassen H a, Cheviron Z a, Freedman AH *et al.* (2010) Spatial modelling and landscape-level approaches for visualizing intra-specific variation. *Molecular Ecology*, **19**, 3532–48.
Tinner W, Theurillat J (2003) Uppermost limit, extent, and fluctuations of the timberline and treeline ecocline in the Swiss Central Alps during the past 11,500 years. *Arctic, Antarctic, and Alpine Research*, **35**, 158–169.
Veith M, Kosuch J, Vences M (2003) Climatic oscillations triggered post-Messinian speciation of Western Palearctic brown frogs (Amphibia, Ranidae). *Molecular phylogenetics and evolution*, **26**, 310–327.
Veith M, Vences M, Vieites DR, Nieto-roman S, Palanca A (2002) Genetic differentiation and population structure within Spanish common frogs (Rana temporaria complex; Ranidae, Amphibia). *Folia Zoologica*, **51**, 307–318.
Vences M, Hauswaldt JS, Steinfartz S *et al.* (2013) Radically different phylogeographies and patterns of genetic variation in two European brown frogs, genus Rana. *Molecular phylogenetics and evolution*, **68**, 657–70.
Vergeer P, Kunin WE (2013) Adaptation at range margins: common garden trials and the performance of Arabidopsis lyrata across its northwestern European range. *New Phytologist*, **197**, 989–1001.
Vitti JJ, Grossman SR, Sabeti PC (2013) Detecting Natural Selection in Genomic Data. *Annu. Rev. Genet*, **47**, 97–120.
Wang IJ (2012) Environmental and topographic variables shape genetic structure and effective population sizes in the endagered Yosemite toad. *Diversity and Distributions*, **18**, 1033–1041.
Wang IJ, Glor RE, Losos JB (2013) Quantifying the roles of ecology and geography in spatial genetic divergence. *Ecology letters*, **16**, 175–82.
Weir BS, Cockerham CC (1984) Estimating F-Statistics for the Analysis of Population Structure. *Evolution*, **38**, 1358–1370.
Wigginton JE, Cutler DJ, Abecasis GR (2005) A note on exact tests of Hardy-Weinberg equilibrium. *American Journal of Human Genetics*, **76**, 887–93.

Willis KJ, Whittaker RJ (2008) The refugial debate. *Science*, **287**, 1406–1407.
Yannic G, Basset P, Hausser J (2008) Phylogeography and recolonization of the Swiss Alps by the Valais shrew (Sorex antinorii), inferred with autosomal and sex-specific markers. *Molecular Ecology*, **17**, 4118–4133.
Yannic G, Pellissier L, Dubey S *et al.* (2012) Multiple refugia and barriers explain the phylogeography of the Valais shrew, Sorex antinorii (Mammalia: Soricomorpha). *Biological Journal of the Linnean Society*, **105**, 864–880.
Yeaman S (2015) Local Adaptation by Alleles of Small Effect. *The American Naturalist*, **186**, S74–S89.

Appendix Chapter 3

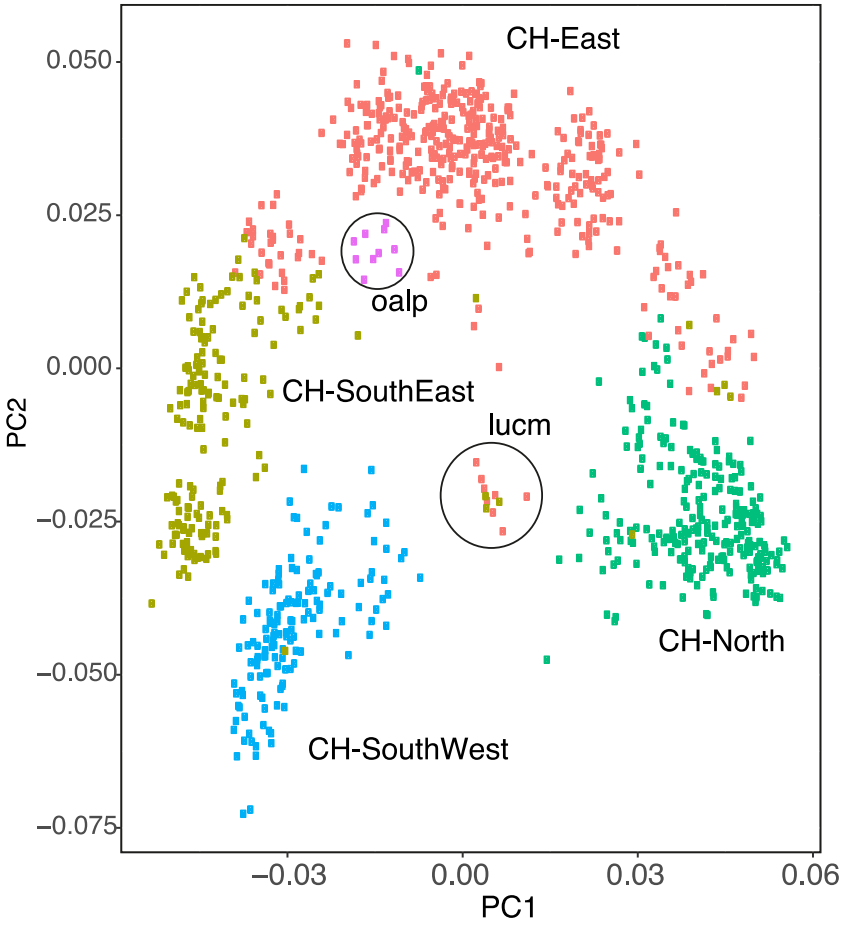


Figure S3.1 The distribution of genomic variation visualised using a Principal Component Analysis (PCA) of all individuals. PC1 explained 19.5% and PC2 9.3% of the genomic variance. Samples are coloured according to their geographic genetic clusters. CHS-SouthEast individuals that cluster with CH-East were from Lukmanierpass (lucm) in central Switzerland. Previous work (Chapter 1) suggest that this might be a pass used for migration between CH-North and CH-South. For a second population, pozz from mid-Ticino, an individual clustered with CHS-SouthWest and a few with CH-SouthEast. Mitochondrial DNA showed evidence that this population was a mix between CH-South and CH-North. One CH-North individual from population meie, near Interlaken, grouped with CH-East. The pink population (oalp) is a geographically-isolated central population that did not fit obviously within any single region.

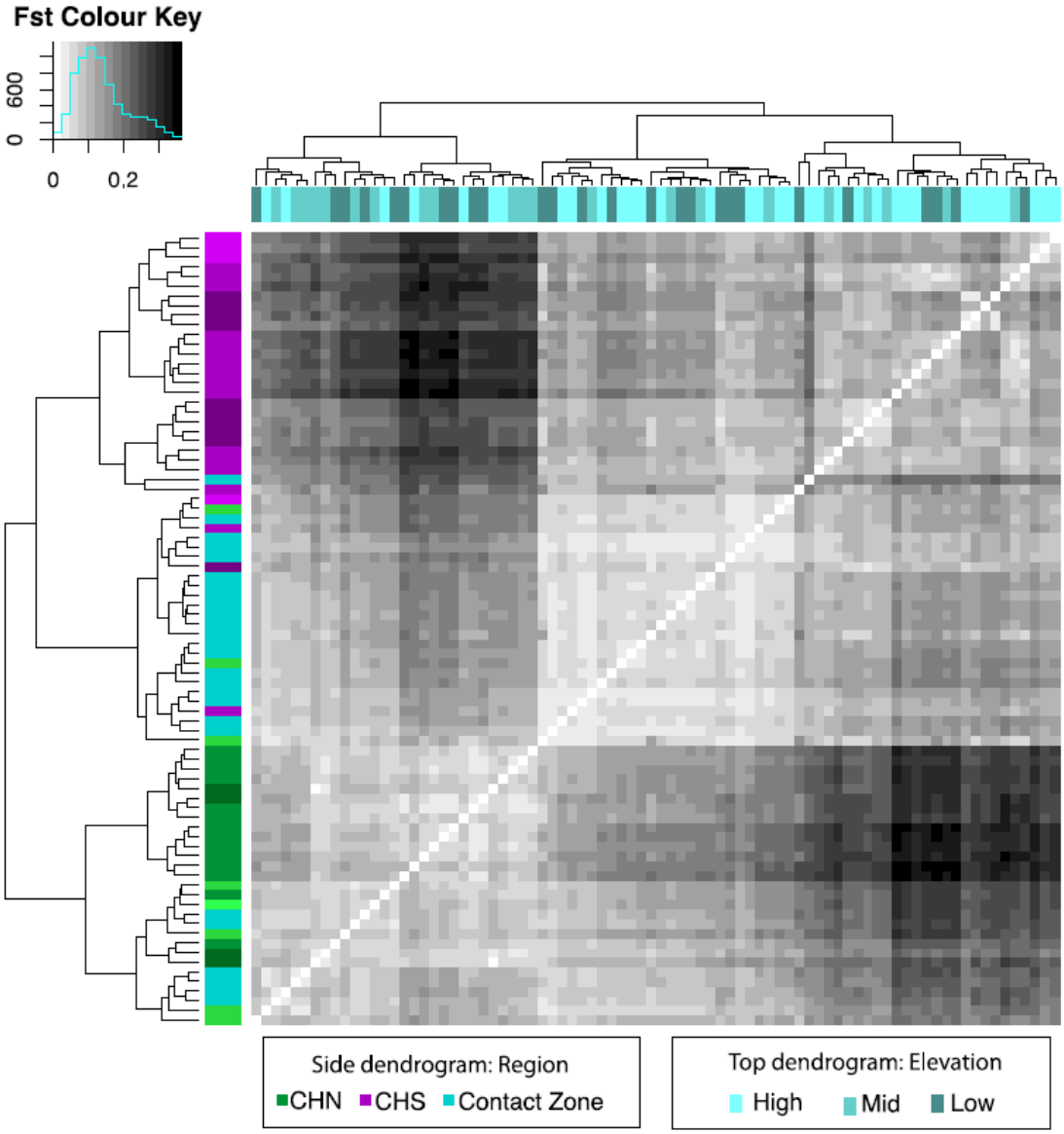


Figure S3.2 Heatmap of the pairwise F_{ST} between all populations, where darker colours indicate higher F_{ST} (refer to F_{ST} colour Key). Unrooted dendrograms to the left and above the heatmap were calculated based on euclidean distances based on F_{ST} matrix. Colours to the left of the heatmap reflect the geographic genetic clusters as identified in Fig. 2. Colours above the heatmap show the elevation of each population (Low < 901 m; Mid 901-1700 m; High >1700 m

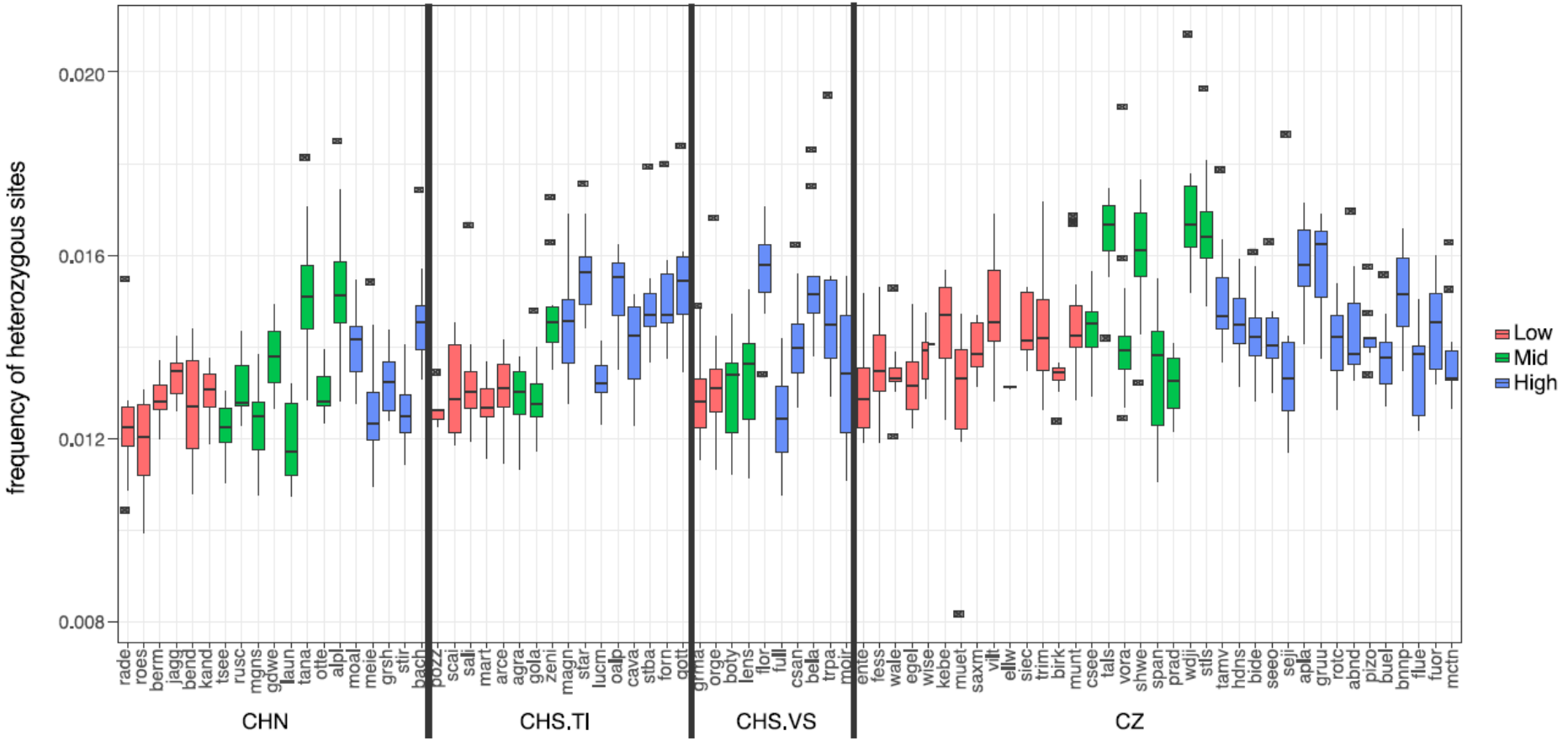


Figure S3.3 Frequency of heterozygous sites across all sequenced sites shown per population. Each bar shows the mean and range of the frequency per population. Bars are coloured according to the site elevation (red/Low <901 m; green/Mid: 901-1700 m; blue/High: >1701 m). Populations are arranged in their geographic geographic regions, with sites sorted from lowest to highest elevation in each region.

Table S3.1 Sample coordinates (Lat & Long), elevation (Elev; in m), and the geographic region (Region) to which they were assigned for each population (Pop). The number of samples (n) analysed and measures of genetic diversity are shown based on genomic (RAD) and mitochondrial (*cytb*) data. F_{IS}: Fixation index, H_S: mean genetic diversity; H_O: mean observed heterozygosity; Hd: haplotype diversity; π: nucleotide diversity. CH-N: CH-North; CH-SE: CH-SouthEast; CH-SW: CH-SouthWest; CH-E: CH-East.

					SNP data				mtDNA data		
Region	**Pop**	**Lat**	**Long**	**Elev**	**n**	**F_{IS}**	**Hs**	**Ho**	**n**	**Hd**	**π**
CH-N	rade	47.027	7.269	450	10	0.097	0.163	0.142	6	0.00	0.000
CH-N	roes	46.903	7.203	538	10	0.083	0.154	0.136	6	0.00	0.000
CH-N	berm	47.015	7.523	550	10	0.111	0.162	0.138	4	0.00	0.000
CH-N	jagg	46.740	8.057	570	10	0.087	0.210	0.184	6	0.00	0.000
CH-N	bend	46.488	6.878	590	20	0.097	0.165	0.145	4	0.00	0.000
CH-N	kand	46.634	7.690	698	18	0.086	0.166	0.148	6	0.00	0.000
CH-N	tsee	46.554	7.747	1151	18	0.048	0.152	0.144	4	0.00	0.000
CH-N	rusc	46.369	7.240	1315	10	0.103	0.180	0.154	4	0.50	0.001
CH-N	mgns	46.248	6.847	1372	19	0.073	0.203	0.184	6	0.00	0.000
CH-N	gdwe	46.641	8.069	1380	10	0.072	0.175	0.157			
CH-N	laun	46.397	7.333	1381	9	0.109	0.174	0.148	6	0.60	0.001
CH-N	tana	46.346	6.840	1408	20	-0.003	0.180	0.182	6	0.00	0.000
CH-N	otte	46.732	7.364	1455	10	0.059	0.164	0.150	4	0.50	0.001
CH-N	alpl	46.935	8.655	1506	10	-0.014	0.171	0.174			
CH-N	moal	46.344	6.797	1817	10	0.071	0.170	0.152			
CH-N	meie	46.587	7.584	1900	19	0.129	0.163	0.139	6	0.73	0.002
CH-N	grsh	46.658	8.104	1929	10	0.076	0.184	0.164	4	0.00	0.000
CH-N	stir	46.535	7.550	2070	10	0.032	0.142	0.136	6	0.53	0.001
CH-N	bach	46.667	8.025	2265	10	0.106	0.169	0.146	4	0.83	0.002
CH-SE	pozz	46.350	8.957	290	5	0.335	0.197	0.106	2	0.00	0.000

CH-SE	scai	45.979	8.927	298	10	0.063	0.171	0.157	4	0.83	0.010
CH-SE	sali	46.255	8.687	331	10	0.105	0.174	0.150	5	0.60	0.001
CH-SE	mart	46.027	8.942	407	7	0.069	0.167	0.150	6	0.73	0.006
CH-SE	arce	46.156	8.746	408	20	0.100	0.182	0.161	6	0.80	0.002
CH-SE	agra	46.033	8.896	930	10	0.075	0.181	0.163	6	0.93	0.008
CH-SE	gola	46.105	8.965	975	16	0.063	0.185	0.172	5	0.00	0.000
CH-SE	zeni	46.518	8.892	1397	10	0.169	0.202	0.161	4	0.50	0.007
CH-SE	magn	46.433	8.681	1840	10	-0.125	0.183	0.213	6	0.53	0.001
CH-SE	star	46.273	8.773	1892	10	0.114	0.204	0.173	6	0.00	0.000
CH-SE	lukm	46.562	8.798	1922	10	0.194	0.258	0.188	6	0.33	0.005
CH-SE	cava	46.359	9.033	2003	10	0.075	0.178	0.161	6	0.33	0.005
CH-SE	stba	46.489	9.170	2059	10	0.049	0.191	0.178	5	0.70	0.010
CH-SE	forn	46.484	8.580	2089	10	0.083	0.165	0.146	6	0.33	0.001
CH-SE	gott	46.560	8.562	2116	10	0.104	0.205	0.177	6	0.60	0.007
CH-SW	grma	46.241	7.010	425	17	0.107	0.222	0.192			
CH-SW	orge	46.234	7.336	643	19	0.062	0.190	0.174			
CH-SW	boty	46.282	7.396	1225	10	0.069	0.199	0.180	5	0.40	0.002
CH-SW	lens	46.292	7.459	1338	18	0.106	0.199	0.172			
CH-SW	flor	46.278	7.289	1950	10	-0.134	0.183	0.214	5	0.70	0.008
CH-SW	full	46.171	7.095	2074	20	0.104	0.182	0.161	6	0.00	0.000
CH-SW	csan	46.323	7.300	2110	20	0.053	0.172	0.162	6	0.73	0.007
CH-SW	bela	46.378	7.978	2179	10	-0.070	0.191	0.210	4	0.00	0.000
CH-SW	trpa	46.288	7.281	2196	10	0.062	0.169	0.156	4	0.00	0.000
CH-SW	moir	46.102	7.573	2553	12	-0.059	0.171	0.185			
CH-E	ente	47.616	8.667	400	8	0.227	0.187	0.128	6	0.33	0.001
CH-E	wale	47.119	9.095	427	8	0.071	0.195	0.176	6	0.33	0.006
CH-E	egel	47.251	9.495	445	18	0.078	0.224	0.203	5	0.90	0.010

CH-E	wise	47.195	9.479	445	8	0.095	0.236	0.205	4	0.50	0.001
CH-E	kebe	47.542	8.781	453	2	-0.448	0.309	0.444	6	0.53	0.001
CH-E	muet	47.454	8.605	460	9	0.209	0.218	0.153	9	0.00	0.000
CH-E	saxm	47.067	9.380	463	8	0.048	0.225	0.209	4	0.50	0.001
CH-E	vilt	47.032	9.446	486	20	0.101	0.225	0.198			
CH-E	ellw	47.041	9.488	491	2	-0.087	0.226	0.228	5	0.60	0.009
CH-E	siec	46.989	9.551	517	10	-0.003	0.189	0.188	5	0.40	0.006
CH-E	trim	46.904	9.544	545	19	0.083	0.225	0.202	4	0.00	0.000
CH-E	birk	47.296	8.812	550	9	-0.015	0.204	0.214	2	1.00	0.002
CH-E	munt	46.728	9.439	648	18	0.115	0.221	0.189			
CH-E	csee	47.047	9.386	1033	17	0.063	0.229	0.211			
CH-E	tals	47.096	9.133	1086	10	0.044	0.194	0.183	5	0.40	0.001
CH-E	vora	47.158	9.378	1123	19	0.068	0.199	0.183	6	0.00	0.000
CH-E	shwe	47.188	9.330	1159	9	-0.009	0.224	0.225			
CH-E	span	47.075	9.131	1430	20	0.071	0.165	0.152	5	0.90	0.012
CH-E	prad	46.784	9.528	1449	7	0.090	0.223	0.195	6	0.80	0.011
CH-E	wdji	46.807	9.717	1638	10	0.024	0.222	0.213	6	0.33	0.005
CH-E	stls	46.965	9.750	1672	10	0.029	0.221	0.211			
CH-E	tamv	47.026	9.347	1810	20	0.081	0.213	0.191	4	0.00	0.000
CH-E	hdns	46.850	9.763	1918	9	0.083	0.223	0.198	6	0.00	0.000
CH-E	bide	47.142	9.404	1983	20	0.024	0.185	0.179	6	0.00	0.000
CH-E	seeo	47.045	9.576	2030	9	0.077	0.222	0.199	6	0.33	0.006
CH-E	seji	46.797	9.731	2090	10	-0.016	0.216	0.220	6	0.60	0.009
CH-E	apla	46.791	9.498	2118	20	-0.086	0.217	0.245	4	0.00	0.000
CH-E	gruu	46.858	9.787	2120	9	0.063	0.225	0.206	4	0.00	0.000
CH-E	rotc	47.014	9.309	2183	10	0.048	0.208	0.196	4	0.50	0.001
CH-E	fess	47.020	9.137	2184	17	0.071	0.208	0.190			

CH-E	abnd	46.427	9.990	2185	10	0.098	0.181	0.160	4	0.00	0.000
CH-E	pizo	46.980	9.422	2209	10	-0.179	0.168	0.204	6	0.33	0.001
CH-E	buel	46.856	9.734	2260	10	0.057	0.223	0.206	6	0.73	0.002
CH-E	bnnp	46.407	10.028	2342	10	0.109	0.163	0.139	3	0.83	0.012
CH-E	flue	46.748	9.952	2388	10	0.027	0.217	0.209	6	0.00	0.000
CH-E	fuor	46.439	9.827	2494	10	0.076	0.182	0.164	5	0.00	0.000
CH-E	mctn	46.482	9.717	2542	10	-0.014	0.218	0.221			
None	oalp	46.654	8.639	1960	10	0.078	0.215	0.191			

Table S3.2 Mean measures of genetic diversity within each region shown for genome-wide (RAD) markers and mitochondrial (*cytb*) data. F_{IS}: Fixation index; H_S: genetic diversity; H_O: observed heterozygosity; Hd: haplotype diversity; π: nucleotide diversity.

Region	**RAD**			***cytb***	
	F_{IS}	**H_S**	**H_O**	**Hd**	**π**
CH-North	0.08	0.17	0.15	0.23	0.001
CH-SouthEast	0.10	0.19	0.16	0.48	0.004
CH-SouthWest	0.03	0.19	0.18	0.31	0.003
CH-East	0.04	0.21	0.20	0.36	0.003

Table S3.3 Proportion of genetic variance explained by geographic distance and elevation between populations. Results were calculated using a redundancy analysis (RDA). The independent effects of elevation and geographic distance were calculated using a partial RDA where the second explanatory variable was kept constant. We show the variance explained by each model as a proportion of the total variance (Partitioned variance), and of the explainable variance (Proportion constrained). Model fit (R2adj) and significance (p) were calculated with an anova. Results are shown for each geographic region, and for genomic (RAD) and mitochondrial (cytb) data.

SNP data

CH-North

Genetic Variation	Partitioned variance	Proportion constrained	R2adj	P
Total variance	0.21			
Full model: dist + elev (constrained variance)	0.03	1.00	0.05	**0.004**
Pure elev(elev|dist)	0.01	0.43	0.02	0.077
pure geog (geog|dist)	0.01	0.41	0.01	0.155
dist + elev	0.01	0.16	0.02	NA

CH-East

Genetic Variation	Partitioned variance	Proportion constrained	R2adj	P
Total variance	0.26			
Full model: dist + elev (constrained variance)	0.04	1.00	0.09	**0.001**
Pure elev(elev|dist)	0.01	0.26	0.01	0.060
pure geog (geog|dist)	0.02	0.48	0.04	**0.002**
dist + elev	0.01	0.27	0.04	NA

CHS-SouthWest

Genetic Variation	Partitioned variance	Proportion constrained	R2adj	P
Total variance	0.22			
Full model: dist + elev (constrained variance)	0.05	1.00	0.04	**0.171**
Pure elev(elev|dist)	0.03	0.53	0.03	0.178
pure geog (geog|dist)	0.03	0.50	0.02	0.248
dist + elev	0.00	-0.03	-0.01	NA

CHS-East

Genetic Variation	Partitioned variance	Proportion constrained	R2adj	P
Total variance	0.29			
Full model: dist + elev (constrained variance)	0.06	1.00	0.05	0.085
Pure elev(elev|dist)	0.02	0.42	0.01	0.276
pure geog (geog|dist)	0.03	0.50	0.03	0.144
dist + elev	0.00	0.08	0.01	NA

cytb

CH-North

Genetic Variation	Partitioned variance	Proportion constrained	R2adj	P
Total variance	0.08			
Full model: dist + elev (constrained variance)	0.03	1.00	0.24	**0.064**
Pure elev(elev|dist)	0.02	0.67	0.19	**0.053**
pure geog (geog|dist)	0.00	0.06	0.00	0.646
dist + elev	0.01	0.27	0.05	NA

CH-East

Genetic Variation	Partitioned variance	Proportion constrained	R2adj	P
Total variance	0.11			
Full model: dist + elev (constrained variance)	0.01	1.00	0.02	0.290
Pure elev(elev|dist)	0.01	1.41	0.00	0.509
pure geog (geog|dist)	0.00	0.23	0.00	0.495
dist + elev	0.00	-0.64	0.02	NA

CHS-SouthWest

Genetic Variation	Partitioned variance	Proportion constrained	R2adj	P
Total variance	0.12			
Full model: dist + elev (constrained variance)	0.02	1.00	0.00	0.665
Pure elev(elev|dist)	0.01	0.58	0.00	0.509
pure geog (geog|dist)	0.00	0.08	0.00	0.921
dist + elev	0.01	0.34	0.00	NA

CHS-SouthEast

Genetic Variation	Partitioned variance	Proportion constrained	R2adj	P
Total variance	0.10			
Full model: dist + elev (constrained variance)	0.01	1.00	0.00	0.396
Pure elev(elev|dist)	0.01	0.51	0.00	0.344
pure geog (geog|dist)	0.00	0.02	0.00	0.888
dist + elev	0.01	0.47	0.00	NA

Adaptation across elevation in *R. temporaria* is determined by population structure and a geographically heterogeneous climate in the Swiss Alps

Adaptation to environment across a species' range often involves polygenic adaptation to multiple environmental variables. An understanding of how heterogenous environments underlie adaptive divergence is a major goal, particularly in light of biodiversity loss due to climate change. *Rana temporaria* occurs across a large (~2000 m) elevational gradient in the Alps and adaptive divergence in several larval life history traits has been established. However, we do not know what environmental variables underlie this divergence, or whether the same variables are important throughout the Alps. I generate genome-wide markers for individuals from 82 populations across their range in Switzerland. The populations were divided into three geographic groups based on previous population structure analyses. I apply a multi-variate method, redundancy analysis, to each region to determine how much genomic variation underlies adaptive divergence, and identify important environmental variables in each region. I find that environmental variables explain up to 61 % of the genomic variation, but that geographic distance explains more variation at the global scale. I also find that different environmental variables are important in different geographic regions: solar radiation and precipitation are important variables south and season length north of the alpine ridge bisecting Switzerland.

Introduction

Geographic variation in climate and environment across a species range commonly results in clines of adaptation as populations evolve in response to local conditions (Freedman *et al.* 2010; Fitzpatrick & Keller 2015; Schweizer *et al.* 2016). However, the importance of an environmental variable may vary across the broad geographic scale of a species range. Understanding the environmental and genetic underpinnings of variation in effects of climate has important implications for predicting a species' adaptive potential, a particularly important goal in light of anthropogenic climate change.

Characterising the geographic distribution of genomic variation is challenging for several reasons. Firstly, environmental distance can co-vary with geographic distance, which makes it difficult to disentangle genetic variation due to neutral population structure from adaptive variation associated with climate gradients (Rellstab *et al.* 2015). Secondly, adaptation to the environment typically involves polygenic adaptation with many loci contributing to the underlying adaptive trait (Yeaman 2015). The small effect size of each contributing locus can make it difficult to distinguish adaptive loci from neutral genetic variation. Finally, adaptation is usually driven by variation in several environmental variables simultaneously, so that examining environmental variables separately can provide an overly-simplistic view of the importance of each variable. Several statistical methods have been developed in the field of landscape genomics to address these issues while investigating the distribution of genomic variation in heterogeneous environments. Most of these approaches find correlations between a single locus and an environmental variable (Günther & Coop 2013; Frichot *et al.* 2015). However, multivariate approaches are increasingly utilised since they can account for polygenic adaptation to complex environments (e.g. Lasky *et al.* 2012; Wang *et al.* 2013; Harrisson *et al.* 2017).

Future changes in temperature and precipitation are predicted to have a particularly large effect on the abundance and distributions of amphibians in montane regions (McCain & Colwell 2011). We assess the relevance of this

prediction on the European common frog, *R. temporaria*, by examining how much genomic variation is associated with climate across their elevational range in the Alps. *Rana temporaria* occurs across a wide range of habitats throughout Europe. Local adaptation in response to several environmental variables, including season length and temperature during development, has been well documented across the latitudinal and elevational distribution of the species (Merilä *et al.* 2000; Laugen *et al.* 2003a; Palo *et al.* 2003b; Muir *et al.* 2014b). Our goal here was to understand how climate has shaped the geographic distribution of adaptive variation across elevation in *R. temporaria*. To address our questions, we applied a series of multivariate regression analyses using climate predictor variables and the minor allele frequency of genome wide markers as response variables. Specifically we ask: 1) How much genomic variation is associated with adaptation to climate and with geographic structure?, 2) which environmental variables are important drivers of adaptation?, and 3) are the same environmental variables important throughout the range in Switzerland?

Materials and Methods

Sampling

To investigate the geographic distribution of genomic variation in Switzerland, we sampled 82 populations across ~2300 m of elevation in the Swiss Alps during the 2013 breeding season (Fig. 4.1). Approximately 10 eggs from each of 20-30 freshly laid clutches were collected from each site. The eggs were transported to the University of Zürich, where they were hatched in separate water-filled containers in a water bath. The tadpoles were raised approximately to stage 36 (Gosner, 1960), at which point they were euthanised and stored in ethanol. Each clutch is assumed to represent a single family, thus one tadpole per clutch was used for all work presented here.

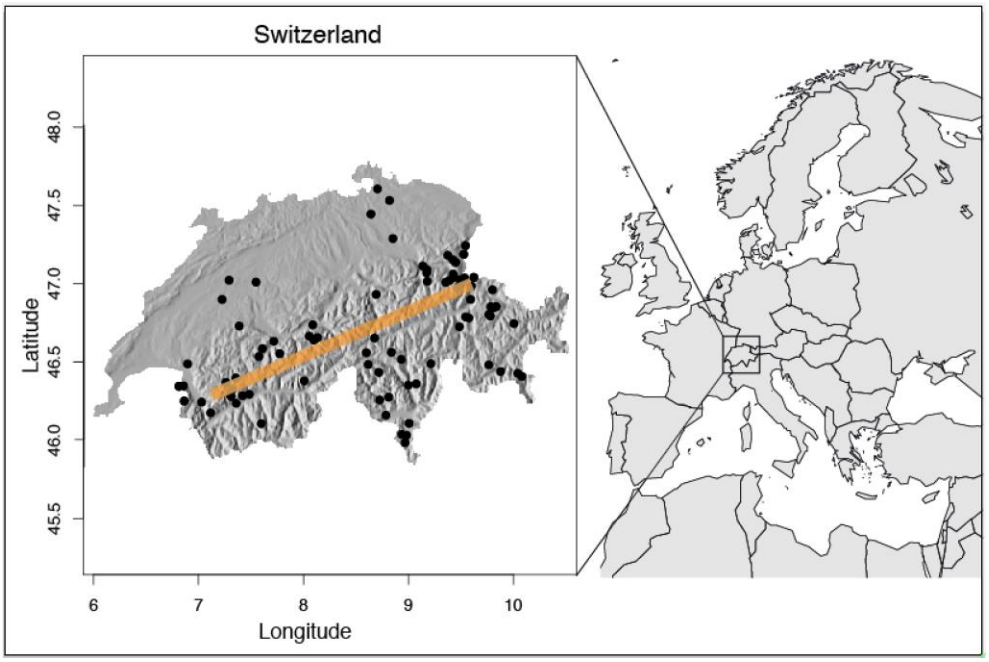


Figure 4.1 Locations of 82 *R. temporaria* populations sampled across their elevational range in Switzerland. The approximate location of the alpine ridge bisecting Switzerland is shown as an orange bar.

Genomic Data - Genome-wide markers were established with a modified version of the double digest restriction-site-associated DNA protocol (Peterson *et al.* 2012). Using *EcoRI* and *MseI* restriction enzymes, we constructed libraries comprising 48 individually barcoded samples each for single-end sequencing (125bp) on an Illumina HiSeq 2500 v4 at the Functional Genomics Center, University of Zurich. Raw sequence reads were demultiplexed using process_radtags (Catchen *et al.* 2011) based on the 5-bp unique barcode + the 4-bp restriction enzyme recognition site, with one mismatch allowed. For the repeated libraries we concatenated the demultiplexed data from both sequencing runs for each individual. We used Trimmomatic v0.33 (Bolger *et al.* 2014) to remove adapter and Illumina-specific sequences from all individuals before further processing.

We used pyRAD (Eaton 2014) for *de novo* assembly and variant calling. This pipeline clusters sequences for each individual separately based on a user specified clustering threshold; that is the percentage of base pairs that should align between sequences for a cluster to be called. Clusters are then aligned between individuals, and only those found in a user specified minimum number of individuals are kept as loci. If the clustering threshold is

set too high, there is a risk of clustering loci together. This would be seen as an artificial increase the number of variants (SNPs), and a decrease in the total number of loci found. However, if the threshold is set too low, clusters that belong to a single locus will be split into two loci, thus decreasing the number of variants found and increasing the number of loci identified (Ilut *et al.* 2014). We tested clustering thresholds of 90-99% on a subset of the data, and chose 94% as the optimum based on the distribution of nucleotide diversity and total number of loci identified. We allowed a Phred quality score of <20 in a maximum of 4 sites per sequence. To avoid including singletons and spurious loci we kept clusters only if they had >5x coverage per individual, and were found in at least 4 individuals. Since the *de novo assembly* and variant calling step is the most time consuming, we applied minimal quality filters here to ensure that we maximise the number of loci identified in the data. Further quality filtering as described in the next step can be implemented and repeated rapidly.

SNP filtering and validation - Possible sequencing errors, uninformative SNPs, paralogs, and missing data were removed from the pyRAD variant file using the following filters: 1) SNPs with a minor allele count of 60 (that is 5.8% in our dataset) were removed, since alleles that occur at a low frequency are likely to be sequencing error, or to bias tests for selection (Roesti *et al.* 2012), 2) we allowed up to 50% missingness, and removed variants using the *--max-missing* function in VCFtools v0.1.12b (Danecek *et al.* 2011). 3) To reduce linkage in the dataset we retained only one SNP per locus using the *--thin* function in VCFtools. 4) All individuals genotyped at less than 55% of the loci were removed using the *--remove* function in VCFtools. 5) We tested whether loci were in Hardy Weinberg Equilibrium (HWE) using PLINK (Purcell *et al.* 2007), and removed any loci deviating from HWE or with an overall observed heterozygosity of more than 0.6.

Climate Data - We obtained mean monthly precipitation, mean monthly global solar radiation, and daily mean monthly maximum and minimum temperatures that had previously been derived for Switzerland based on data

between 1980 and 1989, and are available as 200 m gridded data (Gugerli *et al.* 2008). Elevation at each site was estimated from Google Earth v 10.6.8.

To determine biologically relevant differences between the sampled sites, we derived four variables based on estimates of *R. temporaria* spawning dates and development times at each site. Although the dates measured in any single year are probably affected by weather conditions, the 2013 data are likely to capture the relative difference in timing between sites spanning the elevational gradient. The mean temperature on the spawning date was estimated by interpolating the minimum and maximum monthly temperature data.

Development time for *R. temporaria* eggs and tadpoles range from 40 - 80 days, with a mean of 60 days between spawning and the emergence of the first metamorphs across the latitudinal gradient (Laugen *et al.* 2003b). To capture the climate variation during the larval period, we calculated the mean temperature, mean solar radiation, and mean precipitation during the 60 days after spawning. The length of the thermal growth season was calculated as the number of days with a mean temperature above 6 °C, the estimated threshold above which *R. temporaria* tadpoles are capable of development (Laurila *et al.* 2001; Laugen *et al.* 2003b; Muir *et al.* 2014a)

Topographic shadow at each site was calculated from a 25 m digital elevation model. We determined the date on which the sun clears the topographic horizon for each of 20 rays extending 5 km southwards from the breeding site, sweeping across a pie slice at evenly-spaced intervals between 135-225 degrees. The duration of topographic shadow was the date at which sun first reached the pond averaged across all 20 rays.

At high elevations, the onset of spawning in *R. temporaria* can be limited by the availability of ice-free water bodies. As a proxy we used a snow model to estimate the Julian calendar day on which the snow melts below 20 cm and 10 cm at each site. The snow model was a statistical model that predicted the depth of snow at each frog population as a function of five variables: elevation, aspect (0-1 scale, N-S), hillshade index for a sun azimuth of due south on 1 May (from Horn 1981; integrates aspect, slope, and angle of the sun), topographic shading (described above), and a weighted average of the measured depth of snow at snow stations within 60 km distance and

500 m elevation. Daily snow depth data from 1995-2016 were obtained from the several hundred snow stations operated by SLF (Swiss Federal Institute for Forest, Snow and landscape Research) and MeteoSchweiz (Federal Office of Meteorology and Climatology) found throughout the Swiss Alps. Snow-station measurements were weighted by a linearly declining function of the geographic distance and difference in elevation between the target site and the snow station. Snow stations above 2500 m were excluded since all sampled sites occur below this elevation. The model parameters were estimated by fitting the model separately for each day of the year, using data from 21 years and 316 snow stations.

To find the minimal set of variables that best describes the climate variation across elevation in Switzerland we calculated the absolute pairwise ranked correlation (Spearman's rho) between the eight biologically relevant climate variables and elevation across all 82 sample sites using the *cor* function in R. We retained variables with correlations of $r < 0.8$ as determined using the *findCorrelation* function in the R package *caret v.6.0-71.*

Five climate variables were selected to represent the environmental gradient (Table 4.1): 1) pcpt.60d: mean daily precipitation during the 60 days after spawning, 2) temp.laying.date: the mean temperature on the estimated spawning date, 3) sol.rad.60d: mean daily solar radiation during the 60 days after spawning, 4) shadow.days: the average Julian date until which a site receives topographic shadow, 5) day10cm: Julian date on which the snow melts below 10 cm. Temp.laying.date was strongly correlated with the mean temperature during development, while day10cm was correlated with elevation, day20cm, and duration of the growing season (days above 6 ° C; Fig. S4.1 & S4.2).

Analyses

Subdivision of data into geographic regions - Based on an analysis of the population structure of *R. temporaria* in Switzerland (Chapter 2), we subdivided the dataset into five geographic datasets. Several studies

investigating signatures of adaptation to environment have shown that there is little overlap in the adaptive loci identified between different regions (e.g. Poncet *et al.* 2010; Rellstab *et al.* 2016). This is because the likelihood of a selection acting on an adaptive locus depends on their starting allele frequencies in the population (Messer & Petrov 2013). We account for that here by dividing the dataset into geogaphic groups identified based on their colonisation history and population structure across Switzerland (Chapter 1 and 2). This subdivision also allows us to compare regional signals of adaptation to those variables that determine genomic divergence at the global level when we analyse the full dataset.

Each of the datasets was filtered to remove low frequency minor alleles (MAC < ~5 %) as these could bias tests for selection. Although a minor allele frequency filter had been applied to the full dataset, subdividing the dataset would result in loci that fall below this MAC threshold within each sub-group. The final datasets used in further analyses were: 1) CH-North (229 individuals from 19 populations, 5265 loci), 2) CH-South (275 individuals from 24 populations, 6339 loci), 3) CH-East (403 individuals from 37 populations, 7288 loci), 4) CH-SouthWest (135 individuals from 10 populations, 5835 loci), 5) CH-SouthEast (140 individuals from 14 populations, 5692 loci).

Table 4.1 Description of all the climate and environmental variables examined in this study. Five least correlated variables were selected to represent the environmental gradient (indicated with bold text and *)

Variable	Description	
lat	latitudinal coordinates of sample site	geographic location
long	longitudinal coordinates of sample site	
elev	elevation above sea level (m)	
min.temp.Apr-Aug	mean monthly minimum temperature (°C) for April - August over 10 years	Climate variables
max.temp.Apr-Aug	mean monthly maximum temperature (°C) for April - August over 10 years	
mean.min.temp	mean annual minimum temperature over 10 years	
mean.max.temp	mean annual maximum temperature over 10 years	
mean.annual.temp	mean temperature (°C) [(mean.min.temp+mean.max.temp)/2]	
pcpt.Apr-Aug	mean monthly precipitation (cm) for April - August over 10 years	
annual.pcpt	mean annual precipitation over 10 years (cm)	
SD.annual.pcpt	standard deviation of mean annual precipitation over 10 years (cm)	
annual.solar.rad	mean annual solar radiation (W/m2) over 10 years	
shadow.days*	day of the year (where 1 = 1 Jan) when site is no longer in topographic shadow	
day10cm*	day of the year (where 1 = 1 Jan) on which snow melts below 10 cm	
day20cm	day of the year (where 1 = 1 Jan) on which snow melts below 20 cm	
days.above.6	mean number of days annually with a mean temperature above 6 °C	Biologically relevant variables
temp.laying.date*	mean temperature on the estimated spawning date (°C)	
pcpt.60d*	total precipitation (cm) during 60 days after spawning	
sol.rad.60d*	total solar radiation (W/m2) during 60 days after spawning	
mean.temp.60d	mean temperature(°C) during 60 days after spawning	

We used redundancy analyses (RDA) to determine the relative contribution of climate and geographic distance on neutral genetic divergence. RDA is a multivariate regression where both the predictor and response variables are multivariate, thus it performs better than linear regressions where co-linearity between predictor variables could obscure relationships (Legendre & Fortin 2010; Legendre & Legendre 2012). RDA finds the maximum variance explained by linear orthogonal combinations of the predictor variables, where each axis is independent.

To determine the genomic variance attributable to geographic distance and climate, we created two predictor matrices corresponding to spatial variables (described below) and the five climate variables previously identified. Spatial patterns between sites were modeled using principal components of neighbourhood matrices (PCNM) obtained from a matrix of great circle distances between sites. This method is used to quantify complex patterns at a range of spatial scales (Borcard & Legendre 2002). The matrix of great circle distances was calculated based on geographic coordinates of each site using the *geosphere* package in R. A distance threshold was calculated as the maximum distance of the minimum spanning tree within a region (i.e. four times the distance distance required to join all sites together, in km). Distances exceeding the threshold were assigned four times the threshold to resolve fine and course scale spatial structure (Borcard & Legendre 2002). The threshold distance for each region was: CHall = 37.8 km, CH-North = 50.4 km, CH-East = 122.3 km, CH-South = 47.7 km, CH-SouthWest = 41.0 km, CH-SouthEast = 19.2 km. Eigenvectors of the distance matrix were calculated using the *pcnm* function in the *vegan* package in R. Significance of each eigenvector was tested using the 999 bootstrap permutations as implemented using the *test.scores* function in *spacemakeR v.0.0.5* in R. Only positive and significant eigenvectors were kept for analysis.

The multivariate response was the minor allele frequency for each SNP, calculated using *vcftools* and Plink v.1.0.7. For improved performance of *pcnm* in modeling complex spatial features (Borcard & Legendre 2002), we detrended the response vector with the *decostand* function using the Hellinger method in the *vegan* package in R. This method removes linear trends in the

data by calculating the square root of each element divided by the sum of elements in that row.

We applied a series of analyses using the RDA framework to investigate the contribution of climate and space to genomic variation in this dataset.

Relative contribution of environment and geography to total genomic variation

The first analysis determined the relative contribution of climate and space to the total genomic variation across the different geographic regions. To determine the total variance explained we ran the full RDA model including all climate and spatial predictors. Next we used a partial RDA (pRDA) where the effect of one set of the predictor variables is removed prior to applying a standard RDA. This allows us to partition the proportion of genetic variance attributable to space and climate as defined by the included variables. The interactive contribution of climate and space was calculated as the remainder of the explainable variance (i.e. RDAfull - (pRDAg + pRDAe), where pRDAfull represents the proportion of variance explained by the model with both geographic and environmental predictors, pRDAg is the proportion of variance explained by the model without environmental predictors, and pRDAe is the proportion of variance explained by the model without geographic predictors. All analyses were conducted using the *vegan v2.4.1* package (Oksanen *et al.* 2015) in R. The overall and residual variance was calculated for each model, and the model significance was tested with a 999 permutations.

Climate variables that contribute the most to genomic variation

Our second aim was to identify the most important climate predictor variables in each geographic region. First we used a stepwise procedure of elimination to select the set of independent predictor variables that best describe the variation in the genomic data. This was done using the *ordistep* function in *vegan* with forward and backward stepwise elimination. This function selects the model with the highest adjusted coefficient of determination (R_{adj}^2). We applied this method to the full RDA model as well as

both partial RDA models within each region. We visualised the results with a triplot of the first three RDA axes of the full RDA model where we include only the significant predictor variables on the plot.

Next we calculated the explanatory contribution of each climate variable as the absolute correlations of that variable to each canonical axis, weighted by the proportion of variance explained by that axis. This was divided by the total variance explained by the full model. This allowed us to rank the importance of environmental variables within each region and compare between regions.

Results

Relative contribution of environment and geography to total genomic variation

Fig. 4.2 and Table S4.3 illustrate the proportion of variance explained by geographic and climate predictor variables in the RDA analysis. The full RDA model explained a large proportion (49 - 86%) of the total genetic variance within every region. At a global scale, the model explained 57% of the overall genetic variation. Most of the variance was captured by the spatial predictors (86%) and very little by climate (9%). Within the four regions, climate and space explained similar proportions of variance apart from CH-SouthWest, for which climate explained about twice as much variation as did space (Fig. 4.2).

Climate variables that contribute the most to genomic variation

Our second aim was to determine the most important climate variables within each region, and compare these between geographic regions. Spatial variables were more important than climate variables, and were significant for the full dataset and at large geographic scales (CH-South, CH-North, and CH-East) (Fig. 4.3, Tables S4.4-S4.6). This result reflects strong the population structure over large geographic scales. Most of the variance was explained by

the difference between northern and southern Switzerland (RDA 1 ~30%), and genetic distance from east to west across the country (RDA 2 ~8%). When assessing the full RDA model for CH-South, two spatial variables were found to be significant. These explained the divergence between CH-SouthWest and CH-SouthEast and some divergence from south to north along each region. Two spatial variables explained most of the divergence in CH-East, with most of the variance (~20%) separating populations along the north-south axis. In CH-North, spatial predictors explained divergence from south west to south east along the alpine ridge, and from the high elevation populations in the southern parts of CH-North to the low elevation populations in the north. At smaller geographic scales in CH-SouthEast and CH-SouthWest most of the variance was explained by climate variables rather than spatial variables.

Despite the large effect of geography on genetic variance, single climate variables explained up to 15 % of the total variance (Table 4.2). Two variables, mean daily precipitation during 60 days after spawning (pcpt.60d) and mean daily solar radiation during 60 days after spawning (sol.rad.60d), were significantly associated with the genomic variance globally, in CH-South, CH-SouthWest, and CH-SouthEast. Pcpt.60d was also the only climate variable significantly associated with genomic variance in CH-East. At larger geographic scales these climate variables explained 1-7% of the total genomic variation, and explained more variance (9-14 %) in CH-SouthWest and CH-SouthEast. Two additional variables (the average Julian date until which a site receives topographic shadow and the temperature on the estimated date of spawning) contributed significantly to the variance in CH-SouthWest. The number of days of topographic shadow explained the divergence between the populations south and north of the main valley in CH-SouthWest and temperature on the date of spawning explained the divergence across elevation. While precipitation and solar radiation were important in the southern regions, the date that snow melted below 10 cm (day10cm) explained 12 % of the genomic variation in CH-North, with no other significant climate variables identified. The explanatory variable of day10cm was roughly aligned along the axis separating high elevation populations in

the eastern parts of CH-North from low elevation populations in the north west of CH-North.

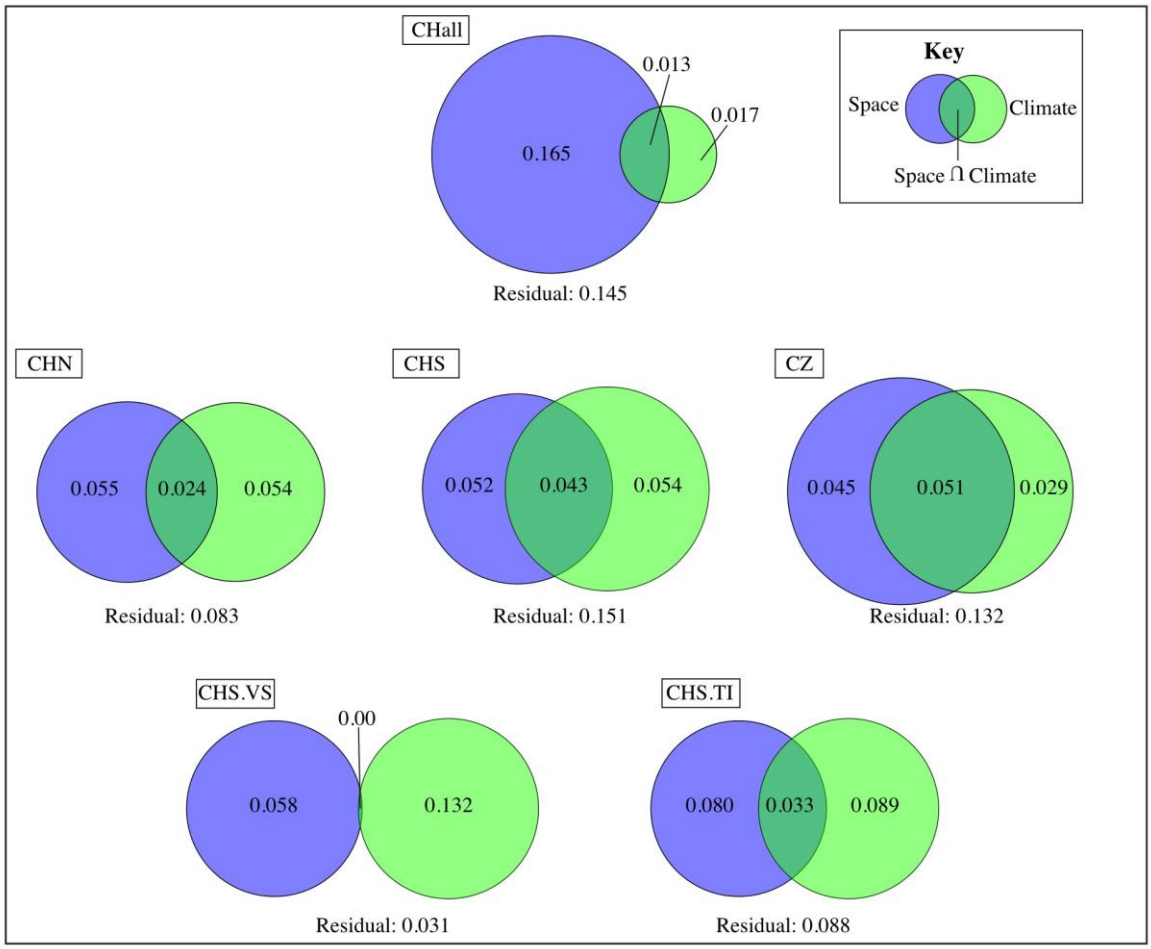
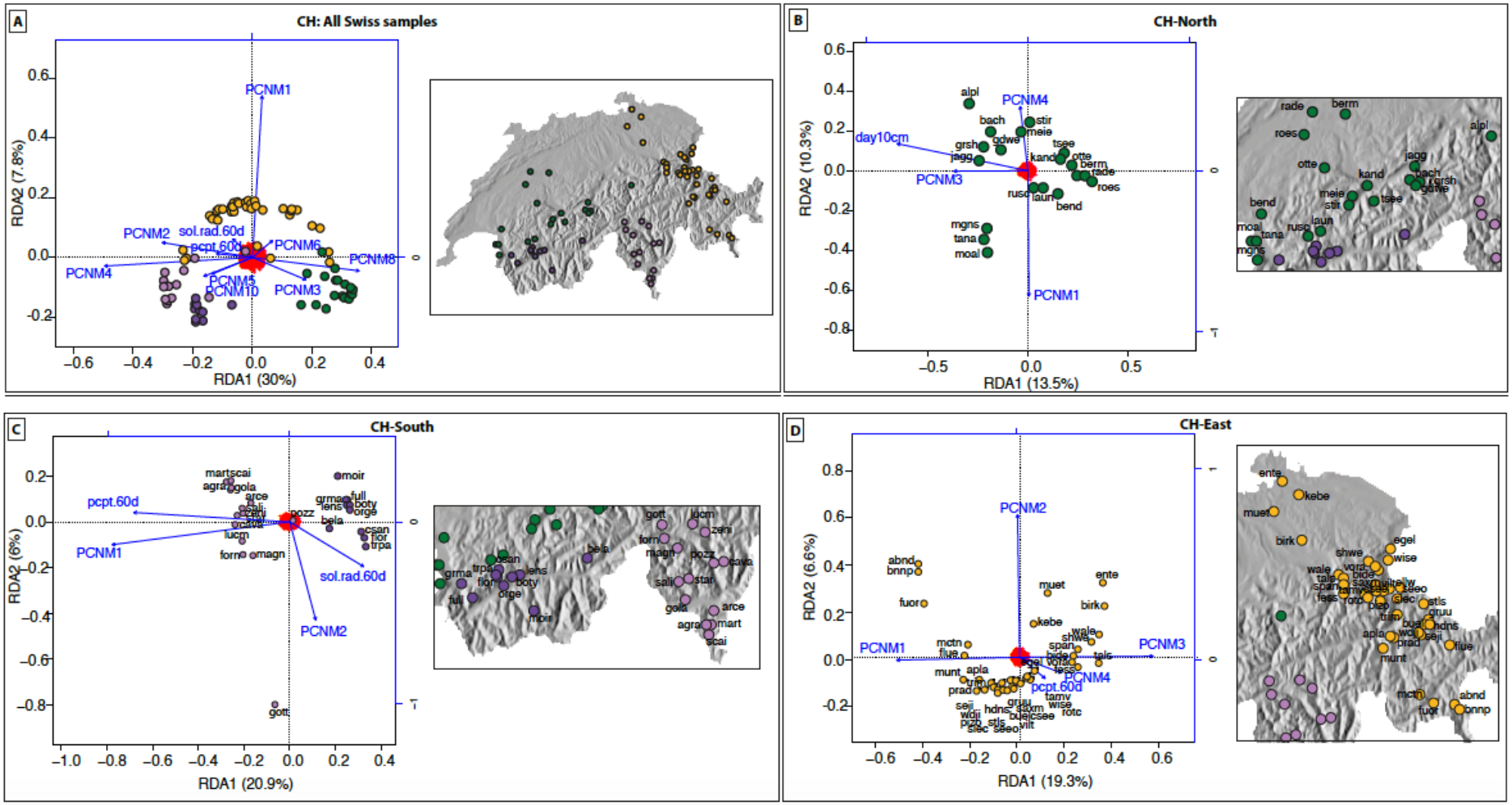


Figure 4.2 Partitioning of genomic variance within each geographic region based on the climate and spatial variables included in each RDA model (See Table S4.3). The residual variance that could not be explained by the full RDA model is shown below each Venn diagram.


A
CH: All Swiss samples
RDA2 (7.8%)
RDA1 (30%)
PCNM1
PCNM2
sol.rad.60d
pcpt.60d
PCNM6
PCNM8
PCNM4
PCNM5
PCNM10
PCNM3
B
CH-North
RDA2 (10.3%)
RDA1 (13.5%)
day10cm
PCNM4
PCNM3
PCNM1
C
CH-South
RDA2 (6%)
RDA1 (20.9%)
pcpt.60d
PCNM1
sol.rad.60d
PCNM2
D
CH-East
RDA2 (6.6%)
RDA1 (19.3%)
PCNM2
PCNM1
PCNM3
PCNM4
pcpt.60d

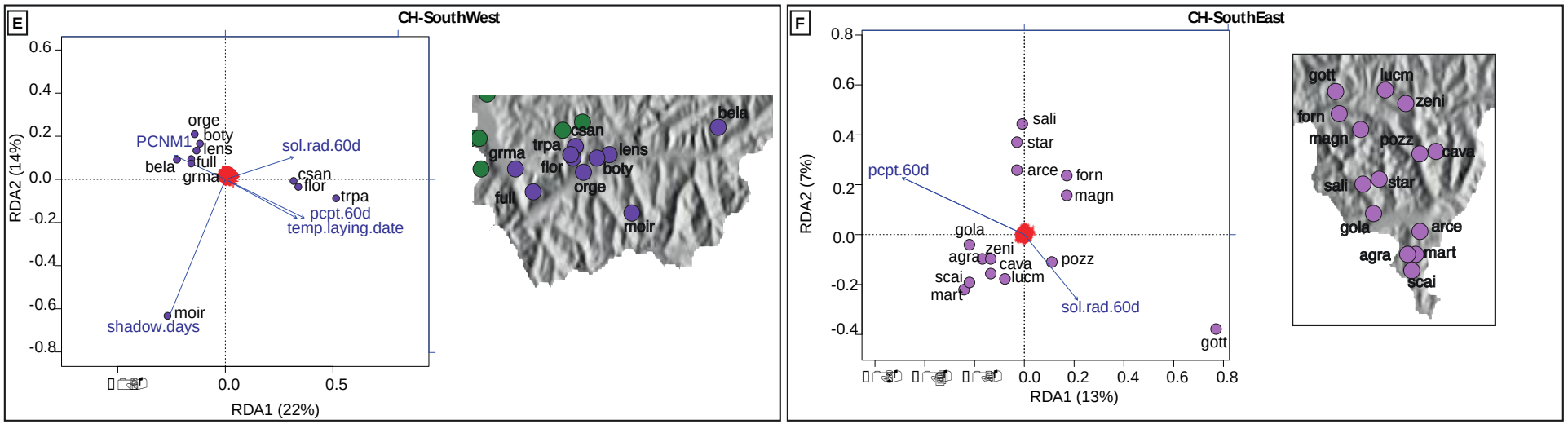


Figure 4.3 Triplots of redundancy analyses (RDA) of each geographic region. RDA included spatial predictor variables (PCNM) and five climate variables (see Table 4.1). The plot was scaled (scale = 3) to maintain the relationship between the points (sample sites) and vectors of the predictor variables. Each triplot shows the location of populations using circles coloured according to their geographic genetic cluster assignment. Population names are shown on the triplot and on the map of each region. Blue arrows on the triplot show the direction and effect size of each predictor variable in describing the genomic variance. Arrows and sample points were plotted using symmetrical scaling to preserve the relationship between all three axes in the plot. Angles between the predictor variables represent the correlation between them. We include only variables found to be significant using a forward selection approach on the full RDA model for each geographic region. Loci are shown as a red "+" in the center of the triplot, where the spread of the symbols represents the SNP variance.

Table 4.2 Proportion of genomic variance that could be explained by the five predictor climate variables globally and in each geographic region. This was calculated as the product of the squared constrained correlation of the variable with an RDA axis and the proportion of variance explained by that RDA axis, summed across all axes. CHall: all samples; CH-N: CH-North; CH-S: CH-South; CH-E: CH-East; CH-SW: CH-SouthWest; CH-SE: CH-SouthEast.

Variable	CHall	CH-N	CH-S	CH-E	CH-SW	CH-SE
sol.rad.60d	0.01		0.07		0.12	0.09
pcpt.60d	0.02		0.02	0.03	0.14	0.11
day10cm		0.12				
temp.laying.date					0.13	
shadow.days					0.15	

Discussion

This study investigates how geographic distance and climate have shaped the geographic distribution of neutral and adaptive genomic variation of *R. temporaria* across their elevational distribution in the European Alps. We applied a series of multivariate analyses to investigate the factors that contribute to the spatial patterns of neutral and adaptive variation globally, and within each of the geographic genetic regions. Although spatial variables explain a large portion of the genomic variance, our models suggest that 5-61 % of genomic variation is associated with variation in climate across the elevational gradient. We find that precipitation and solar radiation during larval development are important south of the Alps, while a measure of snow melting in the spring is the most important north of the Alps. We discuss the importance of these variables for current and future adaptation to environment in *R. temporaria*.

Relative contribution of environment and geography to total genomic variation

Our first aim was to investigate how much variation is associated with space and climate. The RDA models that included climate and spatial variables explained a large portion (49-86 %) of the total genomic variation. Spatial variables were important at global (CHall) and regional geographic scales (individual geographic regions). This may reflect population structure driven by IBD, since the major spatial axes of differentiation reflect the known population structure (Fig. 4.3). In addition, spatial variables had little explanatory power at smaller geographic scales (CH-SouthWest and CH-SouthEast), where IBD is weak (Chapter 2). Spatial variables could also reflect divergence associated with spatially structured environmental variables (Manel *et al.* 2010). Populations from CH-East, for example, occur along a temperature, precipitation, and solar radiation gradient. Similarly, the significant spatial axes explaining divergence in CH-South explain population structure and climatic differences between CH-SouthWest and CH-SouthEast. In both cases IBD is confounded with isolation by environment. Partial RDA models that accounted for spatial structure were a poor fit for the data in all cases except CH-SouthWest (R^2adj 0.322). This suggests that isolation by environment is not the most suitable model for explaining genomic divergence of *R. temporaria* at these larger spatial scales. Overall we found that IBD explains most of the variance globally, while climate and IBD explain similar amounts of variance at regional scales, and climate was more important at fine geographic scales.

Climate variables that contribute the most to genomic variation

We applied a multivariate approach to investigate the correlation between genomic variation and multiple climate variables across Switzerland. This approach is particularly useful when investigating adaptation across heterogeneous environments, since most adaptive processes are expected to be under polygenic control (Kawecki & Ebert 2004). Here we show that precipitation and solar radiation were the most important variables south of

the alpine ridge, while an estimate of the rate of snow melt in the spring was the most important north of the alpine ridge. At smaller geographic scales where IBD was less important, precipitation and solar radiation together explained 20-26% of the total genomic variation. The topographic complexity of Switzerland has resulted in regional differences in climate across the country. Generally, total precipitation levels are higher and total solar radiation lower north of the alpine ridge and in CH-SouthEast than in CH-SouthWest. Within regions, differences in aspect and elevation also result in complex variation in climate at much smaller geographic scales. An example is the gradient in total annual precipitation in CH-SouthEast, which increases from the south east to the north west. Compared to similar studies (Lasky *et al.* 2012; Hecht *et al.* 2015), the our results results imply that there is geographically fine-scale adaptation in response to the heterogeneous environment in the Alps.

A closer examination of the RDA results provides evidence that specific environmental drivers underlie divergence at much smaller geographic scales. Here we explore particular examples in CH-SouthWest (Fig. 4.3E). Two additional climate variables explained a large portion (28%) of the genomic variance in CH-SouthWest. These were the number of days that a pond experiences topographic shadow (shadow.days) and the estimated temperature on the spawning date (temp.laying.date). Shadow.days mainly explained the difference between the single population sampled on the north facing side of the main CH-SouthWest valley (moir) and the rest of the populations (Fig. 4.3F). This central valley in CH-SouthWest is heavily developed and probably creates a barrier to gene flow between populations; thus, we expect genetic divergence between these populations to be higher than estimated from IBD alone. This implies that the effect size of shadow.days (15%) is an overestimate. Nevertheless, topographic shadow would slow the rate that snow melts at these sites, so we expect adaptation to a shorter development season in these populations. A combination of temp.laying.date and precipitation explained the divergence between three high elevation CH-SouthWest populations from the rest (csan, trpa, flor; Fig. 4.3 F). These populations are located close to the alpine ridgeline demarking the geographic separation between the CH-North and CH-South genetic

clusters, and may be at the edge of their dispersal range along this elevational gradient. Interestingly, the environment of the next closest high elevation population (full) was comparable to the two diverged high elevation populations (csan & trpa), but did not group with them in the RDA. This suggests differences in the genomics of adaptation to temperature and precipitation across elevational gradients despite close geographic proximity. The three diverged populations (csan, trpa, and flor; Fig. 4.3E) are genetically more similar; they are located 1.2 - 5 km apart with relatively low genetic divergence (F_{ST} = 0.05) between them, while full is 18-22 km away (F_{ST} = 0.1). All four populations received ~4 mm more precipitation during development, while temp.laying.date was 1.5-2 °C higher for all but one of the diverged populations. Fine scale sampling of the elevational regions is needed to investigate this further.

Adaptation to precipitation

Like many amphibians, *R. temporaria* requires standing water for the survival of eggs and larvae. In southern Switzerland, precipitation during larval development explains a large proportion of the genomic variation in both CH-SouthWest (14%) and CH-SouthEast (11%). Previous studies have shown that amphibians are more threatened by decreasing precipitation than by increasing temperature under future climate scenarios (e.g. Pounds & Crump 1994; Stewart 1995; pounds et al. 1999). This can be attributed to a decrease in the number of suitable ponds, or desiccation of ponds after spawning. *Rana temporaria* tend to spawn in shallow water near the edge of permanent or ephemeral ponds, and clutches are often vulnerable to changes in water levels. A decrease in water levels can expose the upper eggs in a clutch, which leads to damage and loss of eggs through desiccation or freezing. Pond desiccation after the eggs have hatched can lead to higher tadpole densities, which results in metamorphs that emerge earlier and are smaller in size (Brady & Griffiths 2000; Laurila *et al.* 2002; Loman & Claesson 2003). This has subsequent negative effects on juvenile survival and adult fitness. Future climate scenarios predict a mean decrease in precipitation under all scenarios, and more summer drying spells, particularly in the south and west

of the Alps (CH2014–Impacts 2014). This could have a serious impact on *R. temporaria* populations, particularly at low elevations that currently receive the least precipitation. These populations are also the most vulnerable to population fragmentation and habitat destruction due to land-use change (Hitchings & Beebee 1997), which could limit their adaptive potential under changing environmental conditions.

Adaptation to solar radiation

Amphibians are ectothermic and thus rely on solar radiation as a source of heat and energy. Our results suggest that *R. temporaria* have adaptively diverged in response to variation in solar radiation, with a large portion of genomic variation associated with total solar radiation during the 60 days after spawning in CH-SouthWest (12 %) and CH-SouthEast (9 %). Solar energy is particularly important during the egg and larval life stages of *R. temporaria,* where increased water temperatures promote rapid growth and development (Merilä *et al.* 2000; Ståhlberg *et al.* 2001; Laugen *et al.* 2002). Eggs are deposited in shallow water where solar energy can increase the temperature, and tadpoles often bask in very shallow water at the edge of the pond. This behaviour may be particularly important in facilitating development at high elevation sites where the season length is short, but it also increases exposure to ultraviolet radiation (UVR). The harmful effects of UVR on amphibians have been well documented (Croteau *et al.* 2008). There is evidence that *R. temporaria* are adapted to higher levels of UV-B at high elevations (Marquis & Miaud 2008; Marquis *et al.* 2008). However, high levels of UVR exposure in *R. temporaria* has been shown to decrease development rate (Pahkala *et al.* 2001, 2003), reduce hatchling and metamorph size (Pahkala *et al.* 2000, 2002), and lead to morphological abnormalities (Pahkala *et al.* 2001). Levels of UVR increase with the angle of the sun, level of cloud cover, surface reflectance (e.g. from snow), and how thin the atmosphere is, thus higher levels are expected at high latitude and elevation. Future levels of UVR are predicted to increase, particularly in southern Switzerland. Combined

with the predicted decrease in precipitation, this could pose a severe threat to *R. temporaria* populations in southern Switzerland (Alton & Franklin 2017).

Adaptation to season length

The most important environmental variable north of the Alps was a measure of the rate of snow melt in the spring (day10cm), which is highly correlated with elevation across all regions. At high elevations cold temperatures and persistent snow can delay the thawing of ponds and other suitable water bodies until mid summer, thus delaying spawning in these populations. In fact, season length is considerably shorter at high elevations, with as few as 46 days above 6 °C compared to 260 days at low elevations. Season length is an important driver of adaptive divergence in *R. temporaria* across latitude (Laugen *et al.* 2003b; Palo *et al.* 2003b), and is likely to have a similar influence in Switzerland.

Conclusion

Using a multivariate landscape genomic approach, we show that regional differences in climate combined with differences in genetic background and connectivity across regional climate gradients has resulted in differences in adaptive genomic divergence across the European Alps. We conclude that precipitation and solar radiation are the most important climate variables driving adaptation in *R. temporaria* within the Alps, since these were important when Switzerland was examined as a whole. However, the geographic scale of examination and local variation in climate results in regional differences in the importance of climate variables. In Switzerland specifically, solar radiation and precipitation are important variables south and season length north of the alpine ridge bisecting Switzerland.

References

Akaike H (1973) Information theory and an extension of the maximum likelihood principle. In: *2nd International Symposium on Information Theory, Tsahkadsor, Armenia, USSR*, pp. 267–281. Budapest:Akadémiai Kiadó.

Alberto F, Niort J, Derory J *et al.* (2010) Population differentiation of sessile oak at the altitudinal front of migration in the French Pyrenees. *Molecular Ecology*, **19**, 2626–2639.

Alton LA, Franklin CE (2017) Drivers of amphibian declines: effects of ultraviolet radiation and interactions with other environmental factors. *Climate Change Responses*, **4**, DOI 10.1186/s40665-017-0034-7.

Altschup SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic Local Alignment Search Tool. *Journal of Molecular Biology*, **205**, 403–410.

Alvarez N, Thiel-Egenter C, Tribsch A *et al.* (2009) History or ecology? Substrate type as a major driver of spatial genetic structure in Alpine plants. *Ecology Letters*, **12**, 632–640.

Ammann B, Birks HJB, Brooks SJ *et al.* (2000) Quantification of biotic responses to rapid climatic changes around the Younger Dryas — a synthesis. *Paleogeography, Paleoclimatology, Paleoecology*, **159**, 313–347.

Avise JC, Riddle B (2009) Phylogeography : Retrospect and Prospect. *Journal of Biogeography*, **36**, 3–15.

Bachmann JC (2017) Adaptive Divergence across an Elevational Gradient in the Common Frog (Rana temporaria). University of Zurich.

Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, **30**, 2114–2120.

Bonin A (2008) Population genomics: a new generation of genome scans to bridge the gap with functional genomics. *Molecular ecology*, **17**, 3583–4.

Bonin A, Taberlet P, Miaud C, Pompanon F (2006) Explorative genome scan to detect candidate loci for adaptation along a gradient of altitude in the common frog (Rana temporaria). *Molecular biology and evolution*, **23**, 773–83.

Borcard D, Legendre P (2002) All-scale spatial analysis of ecological data by means of principal coordinates of neighbour matrices. *Ecological Modelling*, **153**, 51–68.

Braaker S, Heckel G (2009) Transalpine colonisation and partial phylogeographic erosion by dispersal in the common vole (Microtus arvalis). *Molecular Ecology*, **18**, 2528–2531.

Brady LD, Griffiths RA (2000) Developmental responses to pond desiccation in tadpoles of the British anuran amphibians (Bufo bufo, B. calamita and Rana temporaria). *Journal of Zoology, London*, **252**, 61–69.

Van Buskirk J (2012) Permeability of the landscape matrix between amphibian breeding sites. *Ecology and evolution*, **2**, 3160–7.

Canestrelli D, Cimmaruta R, Costantini V, Nascetti G (2006) Genetic diversity and phylogeography of the Apennine yellow-bellied toad Bombina pachypus, with implications for conservation. *Molecular Ecology*, **15**, 3741–3754.

Canestrelli D, Cimmaruta R, Nascetti G (2008) Population genetic structure and diversity of the Apennine endemic stream frog, Rana italica – insights on the Pleistocene evolutionary history of the Italian peninsular biota. *Molecular Ecology*, **17**, 3856–3872.
Canestrelli D, Nascetti G (2008) Phylogeography of the pool frog Rana (Pelophylax) lessonae in the Italian peninsula and Sicily: Multiple refugia, glacial expansions and nuclear-mitochondrial discordance. *Journal of Biogeography*, **35**, 1923–1936.
Cano JM, Laurila A, Palo J, Merilä J (2004) Population differentiation in G matrix structure due to natural selection in Rana temporaria. *Evolution*, **58**, 2013–2020.
Catchen JM, Amores A, Hohenlohe P, Cresko W, Postlethwait JH (2011) Stacks: Building and Genotyping Loci De Novo From Short-Read Sequences. *G3:Genes, Genomes, Genetics*, **1**, 171–182.
Caye K, Deist TM, Martins H, Michel O, François O (2016) TESS3: Fast inference of spatial population structure and genome scans for selection. *Molecular Ecology Resources*, **16**, 540–548.
CH2014–Impacts (2014) *Toward quantitative scenarios of climate change Impacts in Switzerland*. OCCR, FOEN, MeteoSwiss, C2SM, Agroscope, and ProClim, Bern, Switzerland.
Clement M, Posada D, Crandall KA (2000) TCS : a computer program to estimate gene genealogies. *Molecular Ecology*, **9**, 1657–1659.
Cornetti L, Lemoine M, Hilfiker D *et al.* (2016) Higher genetic diversity on mountain tops: the role of historical and contemporary processes in shaping genetic variation in the bank vole. *Biological Journal of the Linnean Society*, **118**, 233–244.
Croteau MC, Davidson MA, Lean DRS, Trudeau VL (2008) Global Increases in Ultraviolet B Radiation : Potential Impacts on Amphibian Development and Metamorphosis. *Physiological and Biochemical Zoology*, **81**, 743–761.
Crottini A, Andreone F, Kosuch J *et al.* (2007) Fossorial but widespread: the phylogeography of the common spadefoot toad (Pelobates fuscus), and the role of the Po Valley as a major source of genetic variability. *Molecular Ecology*, **16**, 2734–2754.
Danecek P, Auton A, Abecasis G *et al.* (2011) The variant call format and VCFtools. *Bioinformatics*, **27**, 2156–2158.
Dansgaard W, Johnsen SJ, Clausen HB *et al.* (1993) Evidence for general instability of past climate from a 250-kyr ice-core record. *Nature*, **364**, 218–220.
Darnault R, Rolland Y, Braucher R *et al.* (2011) Timing of the last deglaciation revealed by receding glaciers at the Alpine-scale : impact on mountain geomorphology. *Quaternary Science Reviews*, 1–16.
Darriba D, Taboada GL, Doallo R, Posada D (2012) jModelTest 2: more models, new heuristics and parallel computing. *Nature Methods*, **9**, 772.
Debieu M, Tang C, Stich B *et al.* (2013) Co-Variation between seed dormancy, growth rate and flowering time changes with latitude in Arabidopsis thaliana. *PloS one*, **8**, 1–12.
Demesure B, Comps B, Petit RJ (1996) Chloroplast DNA Phylogeography of the Common Beech (Fagus sylvatica L.) in Europe. *Evolution*, **50**, 2515–2520.

Drummond AJ, Suchard MA, Xie D, Rambaut A (2012) Bayesian Phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution*, **29**, 1969–1973.

Dudaniec RY, Spear SF, Richardson JS, Storfer A (2012) Current and historical drivers of landscape genetic structure differ in core and peripheral salamander populations. *PloS one*, **7**.

Eaton DAR (2014) PyRAD: Assembly of de novo RADseq loci for phylogenetic analyses. *Bioinformatics*, **30**, 1844–1849.

Ellis N, Smith SJ, Pitcher CR (2012) Gradient forests : calculating importance gradients on physical predictors. *Ecology*, **93**, 156–168.

Excoffier L, Lischer HEL (2010) An Integrated Software Package for Population Genetics Data Analysis. *Molecular Ecology Resources*, **10**, 564–567.

Fitzpatrick SW, Gerberich JC, Kronenberger JA, Angeloni LM, Funk WC (2015) Locally adapted traits maintained in the face of high gene flow. *Ecology Letters*, **18**, 37–47.

Fitzpatrick MC, Keller SR (2015) Ecological genomics meets community-level modelling of biodiversity: mapping the genomic landscape of current and future environmental adaptation. *Ecology Letters*, **18**, 1–16.

Forester BR, Jones MR, Joost S, Landguth EL, Lasky JR (2016) Detecting spatial genetic signatures of local adaptation in heterogeneous landscapes. *Molecular Ecology*, **25**, 104–120.

Fox J, Weisberg S (2011) *An {R} Companion to Applied Regression*. Thousand Oaks CA: Sage.

Freedman AH, Thomassen HA, Buermann W, Smith TB (2010) Genomic signals of diversification along ecological gradients in a tropical lizard. *Molecular ecology*, **19**, 3773–3788.

Frichot E, François O (2015) LEA : An R package for landscape and ecological association studies (B O'Meara, Ed,). *Methods in Ecology and Evolution*, **6**, 925–929.

Frichot E, Schoville SD, Bouchard G, François O (2013) Testing for associations between loci and environmental gradients using latent factor mixed models. *Molecular Biology and Evolution*, **30**, 1687–99.

Frichot E, Schoville S, de Villemereuil P, Gaggiotti OE, François O (2015) Detecting adaptive evolution based on association with ecological gradients: Orientation matters! *Heredity*, **115**, 22–28.

Garcia VOS, Ivy C, Fu J (2017) Syntopic frogs reveal different patterns of interaction with the landscape: A comparative landscape genetic study of Pelophylax nigromaculatus and Fejervarya limnocharis from central China. *Ecology and Evolution*, **7**, 9294–9306.

Gosner KL (1960) A Simplified Table for Staging Anuran Embryos and Larvae with Notes on Identification. *Herpetologica*, **16**, 183–190.

Goudet J (2005) Hierfstat, a package for R to compute and test hierarchical F-statistics. *Molecular Ecology Notes*, **5**, 184–186.

Gugerli F, Englisch T, Niklfeld H *et al.* (2008) Relationships among levels of biodiversity and the relevance of intraspecific diversity in conservation – a project synopsis. *Perspectives in Plant Ecology, Evolution and Systematics*, **10**, 259–281.

Günther T, Coop G (2013) Robust identification of local adaptation from allele frequencies. *Genetics*, **195**, 205–220.

Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series*, **41**, 95–98.
Harrisson KA, Amish SJ, Pavlova A *et al.* (2017) Signatures of polygenic adaptation associated with climate across the range of a threatened fish species with high genetic connectivity. *Molecular ecology*, **26**, 6253–6269.
Hecht BC, Matala AP, Hess JE, Narum SR (2015) Environmental adaptation in Chinook salmon (Oncorhynchus tshawytscha) throughout their North American range. *Molecular Ecology*, **24**, 5573–5595.
Heiri C, Bugmann H, Tinner W, Heiri O, Lischke H (2006) A model-based reconstruction of Holocene treeline dynamics in the Central Swiss Alps. *Journal of Ecology*, **94**, 206–216.
Hermisson J, Pennings PS (2005) Soft sweeps: molecular population genetics of adaptation from standing genetic variation. *Genetics*, **169**, 2335–52.
Hewitt GM (1996) Some genetic consequences of ice ages, and their role in divergence and speciation. *Biological Journal of the Linnean Society*, **58**, 247–276.
Hewitt G (1999) Post-glacial re-colonization of European biota. *Biological Journal of the Linnean Society*, **68**, 87–112.
Hewitt G (2000) The genetic legacy of the Quaternary ice ages. *Nature*, **405**, 907–913.
Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A (2005) Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, **25**, 1965–1978.
Hitchings SP, Beebee TJC (1997) Genetic substructuring as a result of barriers to gene flow in urban Rana temporaria (common frog) populations: implications for biodiversity conservation. *Heredity*, **79**, 117–127.
Hjernquist MB, Soderman F, Jonsson KI *et al.* (2012) Seasonality determines patterns of growth and age structure over a geographic gradient in an ectothermic vertebrate. *Oecologia*, **170**, 641–649.
Hoban S, Kelley JL, Lotterhos KE *et al.* (2016) Finding the Genomic Basis of Local Adaptation: Pitfalls, Practical Solutions, and Future Directions. *The American Naturalist*, **188**, 000–000.
Hoffmann A a, Sgrò CM (2011) Climate change and evolutionary adaptation. *Nature*, **470**, 479–85.
Huelsenbeck JP, Ronquist F (2001) MRBAYES : Bayesian inference of phylogenetic trees. *Bioinformatics Application Note*, **17**, 754–755.
Ilut DC, Nydam ML, Hare MP (2014) Defining loci in restriction-based reduced representation genomic data from nonmodel species: Sources of bias and diagnostics for optimal clustering. *BioMed Research International*, **2014**, 9 pages.
Jombart T (2008) Adegenet: A R package for the multivariate analysis of genetic markers. *Bioinformatics*, **24**, 1403–1405.
Jombart T, Ahmed I (2011) adegenet 1.3-1: New tools for the analysis of genome-wide SNP data. *Bioinformatics*, **27**, 3070–3071.
Jones MR, Forester BR, Teufel AI *et al.* (2013) Integrating landscape genomics and spatially explicit approaches to detect loci under selection in clinal populations. *Evolution*, **67**, 3455–68.

Kawecki TJ, Ebert D (2004) Conceptual issues in local adaptation. *Ecology Letters*, **7**, 1225–1241.

Lasky JR, Des Marais DL, McKay J *et al.* (2012) Characterizing genomic variation of Arabidopsis thaliana: the roles of geography and climate. *Molecular Ecology*, **21**, 5512–5529.

Laugen AT, Kruuk LEB, Laurila A *et al.* (2005a) Quantitative genetics of larval life-history traits in Rana temporaria in different environmental conditions. *Genetics Research*, **86**, 161–170.

Laugen AT, Laurila A, Jönsson KI, Söderman F, Merilä J (2005b) Do common frogs (*Rana temporaria*) follow Bergmann's rule? *Evolutionary Ecology Research*, **7**, 717–731.

Laugen AT, Laurila A, Merilä J (2002) Maternal and genetic contributions to geographical variation in Rana temporaria larval life-history traits. *Biological Journal fo the Linnean Society*, **76**, 61–70.

Laugen AT, Laurila A, Merilä J (2003a) Latitudinal and temperature-dependent variation in embryonic development and growth in Rana temporaria. *Oecologia*, **135**, 548–554.

Laugen AT, Laurila A, Rasanen K, Merilä J (2003b) Latitudinal countergradient variation in the common frog (Rana temporaria) development rates – evidence for local adaptation. *Journal of Evolutionary Biology*, **16**, 996–1005.

Laurila A, Karttunen S, Meril J (2002) Adaptive Phenotypic Plasticity and Genetics of Larval Life Histories in Two Rana Temporaria Populations. *Evolution*, **56**, 617–627.

Laurila A, Pakkasmaa S, Merila J (2001) Influence of Seasonal Time Constraints on Growth and Development of Common Frog Tadpoles: A Photoperiod Experiment. *Oikos*, **95**, 451–460.

Legendre P, Fortin M-J (2010) Comparison of the Mantel test and alternative approaches for detecting complex multivariate relationships in the spatial analysis of genetic data. *Molecular Ecology Resources*, **10**, 831–844.

Legendre P, Legendre LF (2012) *Numerical ecology*. Elsevier.

Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics Application Note*, **25**, 1451–1452.

Lindgren B, Laurila A (2005) Proximate causes of adaptive growth rates: growth efficiency variation among latitudinal populations of Rana temporaria. *Journal of Evolutionary Biology*, **18**, 820–828.

Lindgren B, Laurila A (2009) Physiological variation along a geographical gradient: is growth rate correlated with routine metabolic rate in Rana temporaria tadpoles? *Biological Journal of the Linnean Society*, **98**, 217–224.

Loman J, Claesson D (2003) Plastic response to pond drying in tadpoles Rana temporaria: tests of cost models. *Evolutionary Ecology Research*, **5**, 179–194.

Lugon-Moulin N, Hausser J (2002) Phylogeographical structure, postglacial recolonization and barriers to gene flow in the distinctive Valais chromosome race of the common shrew (Sorex araneus). *Molecular Ecology*, **11**, 785–794.

Luikart G, England PR, Tallmon D, Jordan S, Taberlet P (2003) The power and promise of population genomics: from genotyping to genome typing.

Nature Reviews Genetics, **4**, 981–94.
Luu K, Bazin E, Blum MGB (2016) pcadapt: An R package to perform genome scans for selection based on principal component analysis. *Molecular Ecology Resources*, **33**, 67–77.
Manel S, Perrier C, Pratlong M *et al.* (2016) Genomic resources and their influence on the detection of the signal of positive selection in genome scans. *Molecular ecology*, **25**, 170–184.
Manel S, Poncet BN, Legendre P, Gugerli F, Holderegger R (2010) Common factors drive adaptive genetic variation at different spatial scales in Arabis alpina. *Molecular ecology*, **19**, 3824–35.
Marquis O, Miaud C (2008) Variation in UV sensitivity among common frog Rana temporaria populations along an altitudinal gradient. *Zoology*, **111**, 309–317.
Marquis O, Miaud C, Lena J-P (2008) Developmental responses to UV-B radiation in common frog Rana temporaria embryos from along an altitudinal gradient. *Population Ecology*, **50**, 123–130.
Mátyás G, Sperisen C (2001) Chloroplast DNA polymorphisms provide evidence for postglacial re-colonisation of oaks (Quercus spp.) across the Swiss Alps. *Theoretical and Applied Genetics*, **102**, 12–20.
McCain CM, Colwell RK (2011) Assessing the threat to montane biodiversity from discordant shifts in temperature and precipitation in a changing climate. *Ecology Letters*, **14**, 1236–1245.
Merila J, Laurila A, Laugen AT, Rasanen K, Pahkala M (2000) Plasticity in age and size at metamorphosis in Rana temporaria: comparison of high and low latitude populations. *Ecography*, **23**, 457–465.
Merilä J, Laurila A, Laugen AT, Räsänen K, Pahkala M (2000) Plasticity in Age and Size at Metamorphosis in Rana Temporaria: Comparison of High and Low Latitude Populations. *Ecography*, **23**, 457–465.
Messer PW, Petrov DA (2013) Population genomics of rapid adaptation by soft selective sweeps. *Trends in Ecology and Evolution*, **28**, 659–669.
Miaud C, Guyetant R, Elmberg J (1999) Variations in life-history traits in the common frog Rana temporaria (Amphibia: Anura): a literature review and new data from the French Alps. *Journal of Zoology*, **249**, 61–73.
Miaud C, Merilä J (2001) Local adaptation or environmental induction? Causes of population differentiation in alpine amphibians. *Biota*, **2**, 31–50.
Muir AP, Biek R, Mable BK (2014a) Behavioural and physiological adaptations to low-temperature environments in the common frog, Rana temporaria. *BMC Evolutionary Biology*, **14**, 1471–2148.
Muir A, Piek R, Thomas R, Mable B (2014b) Local adaptation with high gene flow: temperature parameters drive adaptation to altitude in the common frog (Rana temporaria). *Molecular ecology*, **23**, 561–574.
Nychka D, Furrer R, Paige J, Sain S (2015) fields: Tools for spatial data.
Oksanen J, Blanchet FG, Kindt R *et al.* (2015) Vegan: community ecology package. R package vegan, vers. 2.2-1.
Pahkala M, Laurila A, Meril J (2002) Effects of ultraviolet-B radiation on common frog Rana temporaria embryos from along a latitudinal gradient. *Oecologia*, **133**, 458–465.
Pahkala M, Laurila A, Merila J (2000) Ambient Ultraviolet-B radiation reduces hatchling size in the common frog Rana temporaria. *Ecography*, **23**, 531–

538.
Pahkala M, Laurila A, Merila J (2001) Carry-over effects of ultraviolet-B radiation on larval tness in Rana temporaria. *Proceedings of the Royal Society of London B: Biological Sciences*, **268**, 1699–1706.
Pahkala M, Merila J, Ots I, Laurila A (2003) Effects of ultraviolet-B radiation on metamorphic traits in the common frog Rana temporaria. *Journal of the Zoological Society of London*, **259**, 57–62.
Palo JU, O'Hara RB, Laugen AT *et al.* (2003a) Latitudinal divergence of common frog (Rana temporaria) life history traits by natural selection: evidence from a comparison of molecular and quantitative genetic data. *Molecular ecology*, **12**, 1963–1978.
Palo JU, O'Hara RB, Laugen AT *et al.* (2003b) Latitudinal divergence of the common frog (Rana temporaria) life history traits by natural selection: evidence from a comparison of molecular quantitative genetic data. *Molecular Ecology*, **12**, 1963–1978.
Palo JU, Schmeller DS, Laurila A *et al.* (2004) High degree of population subdivision in a widespread amphibian. *Molecular Ecology*, **13**, 2631–2644.
Parisod C (2008) Postglacial recolonisation of plants in the western Alps of Switzerland. *Botanica Helvetica*, **118**, 1–12.
Parmesan C (2006) Ecological and Evolutionary Responses to Recent Climate Change. *Annual Review of Ecology, Evolution, and Systematics*, **37**, 637–669.
Parmesan C, Yohe G (2003) A globally coherent fingerprint of climate change impacts across natural systems. *Nature*, **421**, 37–42.
Pennings PS, Hermisson J (2006) Soft sweeps II - Molecular population genetics of adaptation from recurrent mutation or migration. *Molecular Biology and Evolution*, **23**, 1076–1084.
Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE (2012) Double Digest RADseq: An Inexpensive Method for De Novo SNP Discovery and Genotyping in Model and Non-Model Species. *PloS one*, **7**.
Polechová J, Barton NH (2015) Limits to adaptation along environmental gradients. *Proceedings of the National Academy of Sciences of the United States of America*, **112**, 6401–6406.
Poncet BN, Herrmann D, Gugerli F *et al.* (2010) Tracking genes of ecological relevance using a genome scan in two independent regional population samples of Arabis alpina. *Molecular ecology*, **19**, 2896–907.
Pritchard JK, Di Rienzo A (2010) Adaptation - not by sweeps alone. *Nature Reviews Genetics*, **11**, 665–7.
Purcell S, Neale B, Todd-Brown K *et al.* (2007) PLINK: A tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics*, **81**, 559–575.
Raj A, Stephens M, Pritchard JK (2014) FastSTRUCTURE: Variational inference of population structure in large SNP data sets. *Genetics*, **197**, 573–589.
Rehm EM, Olivas P, Stroud J, Feeley KJ (2015) Losing your edge: climate change and the conservation value of range-edge populations. *Ecology and evolution*, **5**, 4315–4326.
Rellstab C, Fischer MC, Zoller S *et al.* (2016) Local adaptation (mostly) remains local: reassessing environmental associations of climate-related

candidate SNPs in Arabidopsis halleri. *Heredity*, 1–9.
Rellstab C, Gugerli F, Eckert AJ, Hancock AM, Holderegger R (2015) A practical guide to environmental association analysis in landscape genomics. *Molecular Ecology*, **24**, 4348–70.
Rigling A, Bigler C, Eilmann B *et al.* (2013) Driving factors of a vegetation shift from Scots pine to pubescent oak in dry Alpine forests. *Global Change Biology*, **19**, 229–240.
Rodrigues N, Betto-Colliard C, Jourdan-Pineau H, Perrin N (2013) Within-population polymorphism of sex-determination systems in the common frog (Rana temporaria). *Journal of evolutionary biology*, **26**, 1569–1577.
Roesti M, Salzburger W, Berner D (2012) Uninformative polymorphisms bias genome scans for signatures of selection. *BMC Evolutionary Biology*, **12**, 94.
Roff DA (1996) The evolution of threshold traits in animals. *The Quarterly Review of Biology*, **71**, 3–35.
Rogers AR, Harpending H (1992) Population Growth Makes Waves in the Distribution of Pairwise Genetic Differences. *Molecular biology and evolution*, **9**, 552–569.
Rogivue A, Graf R, Parisod C, Holderegger R, Gugerli F (2018) The phylogeographic structure of Arabis alpina in the Alps shows consistent patterns across different types of molecular markers and geographic scales. *Alpine Botany*, **0**, 0.
Rousset F (1997) Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. *Genetics*, **145**, 1219–1228.
Roy K, Valentine JW, Jablonski D, Kidwell SM (1996) Scales of climatic variability and time averaging in Pleistocene biotas: implications for ecology and evolution. *Trends in Ecology & Evolution*, **11**, 458–463.
Schmeller DS, Palo JU, Merilä J (2008) A contact zone between two distinct Rana temporaria lineages in northern Germany. *Alytes*, **25**, 93–98.
Schneider S, Excoffier L (1999) Estimation of Past Demographic Parameters From the Distribution of Pairwise Differences When the Mutation Rates Vary Among Sites: Application to Human Mitochondrial DNA. *Genetics*, **152**, 1079–1089.
Schweizer RM, VonHoldt BM, Harrigan R *et al.* (2016) Genetic subdivision and candidate genes under selection in North American grey wolves. *Molecular Ecology*, **25**, 380–402.
Sillero N, Campos J, Bonardi A *et al.* (2014) Updated distribution and biogeography of amphibians and reptiles of Europe. *Amphibia-Reptilia*, **35**, 1–31.
Sork VL, Aitken SN, Dyer RJ *et al.* (2013) Putting the landscape into the genomics of trees: approaches for understanding local adaptation and population responses to changing climate. *Tree Genetics & Genomes*, **9**, 901–911.
Ståhlberg F, Olsson M, Uller T (2001) Population divergence of developmental thermal optima in Swedish common frogs, Rana temporaria. *Journal of Evolutionary Biology*, **14**, 755–762.
Stamatakis A (2006) RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics Application Note*, **22**, 2688–2690.
Stapley J, Reger J, Feulner PGD *et al.* (2010) Adaptation genomics: the next

generation. *Trends in ecology & evolution*, **25**, 705–12.
Stefani F, Gentilli A, Sacchi R *et al.* (2012) Refugia within refugia as a key to disentangle the genetic pattern of a highly variable species: The case of Rana temporaria Linnaeus, 1758 (Anura, Ranidae). *Molecular Phylogenetics and Evolution*, **65**, 718–726.
Stehlik I, Blattner FR, Holderegger R, Bachmann K (2002) Nunatak survival of the high Alpine plant Eritrichium nanum (L.) Gaudin in the central Alps during the ice ages. *Molecular Ecology*, **11**, 2027–2036.
Taberlet P, Fumagalli L, Wust-Saucy A-G, Cossons J-F (1998) Comparative phylogeography and postglacial colonization routes in Europe. *Molecular Ecolgy*, **7**, 453–464.
Teacher AGF, Garner TWJ, Nichols RA (2009) European phylogeography of the common frog (Rana temporaria): routes of postglacial colonization into the British Isles, and evidence for an Irish glacial refugium. *Heredity*, **102**, 490–6.
Templeton AR, Crandall KA, Sing CF (1992) Cladistic Analysis of Phenotypic Associations With Haplotypes Inferred From Restriction Endonuclease Mapping and DNA Sequence Data. III. Cladogram Estimation. *Genetics*, **132**, 619–633.
Thomassen H a, Cheviron Z a, Freedman AH *et al.* (2010) Spatial modelling and landscape-level approaches for visualizing intra-specific variation. *Molecular Ecology*, **19**, 3532–48.
Tinner W, Theurillat J (2003) Uppermost limit, extent, and fluctuations of the timberline and treeline ecocline in the Swiss Central Alps during the past 11,500 years. *Arctic, Antarctic, and Alpine Research*, **35**, 158–169.
Veith M, Kosuch J, Vences M (2003) Climatic oscillations triggered post-Messinian speciation of Western Palearctic brown frogs (Amphibia, Ranidae). *Molecular phylogenetics and evolution*, **26**, 310–327.
Veith M, Vences M, Vieites DR, Nieto-roman S, Palanca A (2002) Genetic differentiation and population structure within Spanish common frogs (Rana temporaria complex; Ranidae, Amphibia). *Folia Zoologica*, **51**, 307–318.
Vences M, Hauswaldt JS, Steinfartz S *et al.* (2013) Radically different phylogeographies and patterns of genetic variation in two European brown frogs, genus Rana. *Molecular phylogenetics and evolution*, **68**, 657–70.
Vergeer P, Kunin WE (2013) Adaptation at range margins: common garden trials and the performance of Arabidopsis lyrata across its northwestern European range. *New Phytologist*, **197**, 989–1001.
Vitti JJ, Grossman SR, Sabeti PC (2013) Detecting Natural Selection in Genomic Data. *Annu. Rev. Genet*, **47**, 97–120.
Wang IJ (2012) Environmental and topographic variables shape genetic structure and effective population sizes in the endagered Yosemite toad. *Diversity and Distributions*, **18**, 1033–1041.
Wang IJ, Glor RE, Losos JB (2013) Quantifying the roles of ecology and geography in spatial genetic divergence. *Ecology letters*, **16**, 175–82.
Weir BS, Cockerham CC (1984) Estimating F-Statistics for the Analysis of Population Structure. *Evolution*, **38**, 1358–1370.
Wigginton JE, Cutler DJ, Abecasis GR (2005) A note on exact tests of Hardy-Weinberg equilibrium. *American Journal of Human Genetics*, **76**, 887–93.

Willis KJ, Whittaker RJ (2008) The refugial debate. *Science*, **287**, 1406–1407.
Yannic G, Basset P, Hausser J (2008) Phylogeography and recolonization of the Swiss Alps by the Valais shrew (Sorex antinorii), inferred with autosomal and sex-specific markers. *Molecular Ecology*, **17**, 4118–4133.
Yannic G, Pellissier L, Dubey S *et al.* (2012) Multiple refugia and barriers explain the phylogeography of the Valais shrew, Sorex antinorii (Mammalia: Soricomorpha). *Biological Journal of the Linnean Society*, **105**, 864–880.
Yeaman S (2015) Local Adaptation by Alleles of Small Effect. *The American Naturalist*, **186**, S74–S89.

Appendix Chapter 4

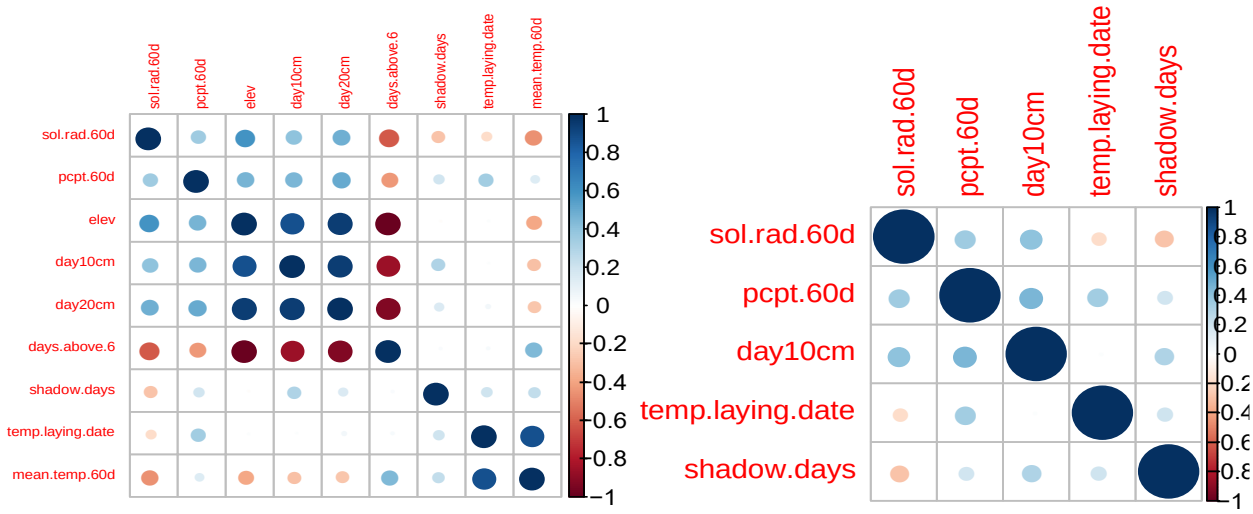


Figure S4.1 Correlation between biologically relevant environmental variables (left), and the five climate variables selected for further analyses (right). Size of the circle is scaled with correlation. A colour scale is used to distinguish positive and negative correlations as shown on the bar to the right of each figure. To reduce correlation between variables we removed those that were most correlated throughout the dataset.

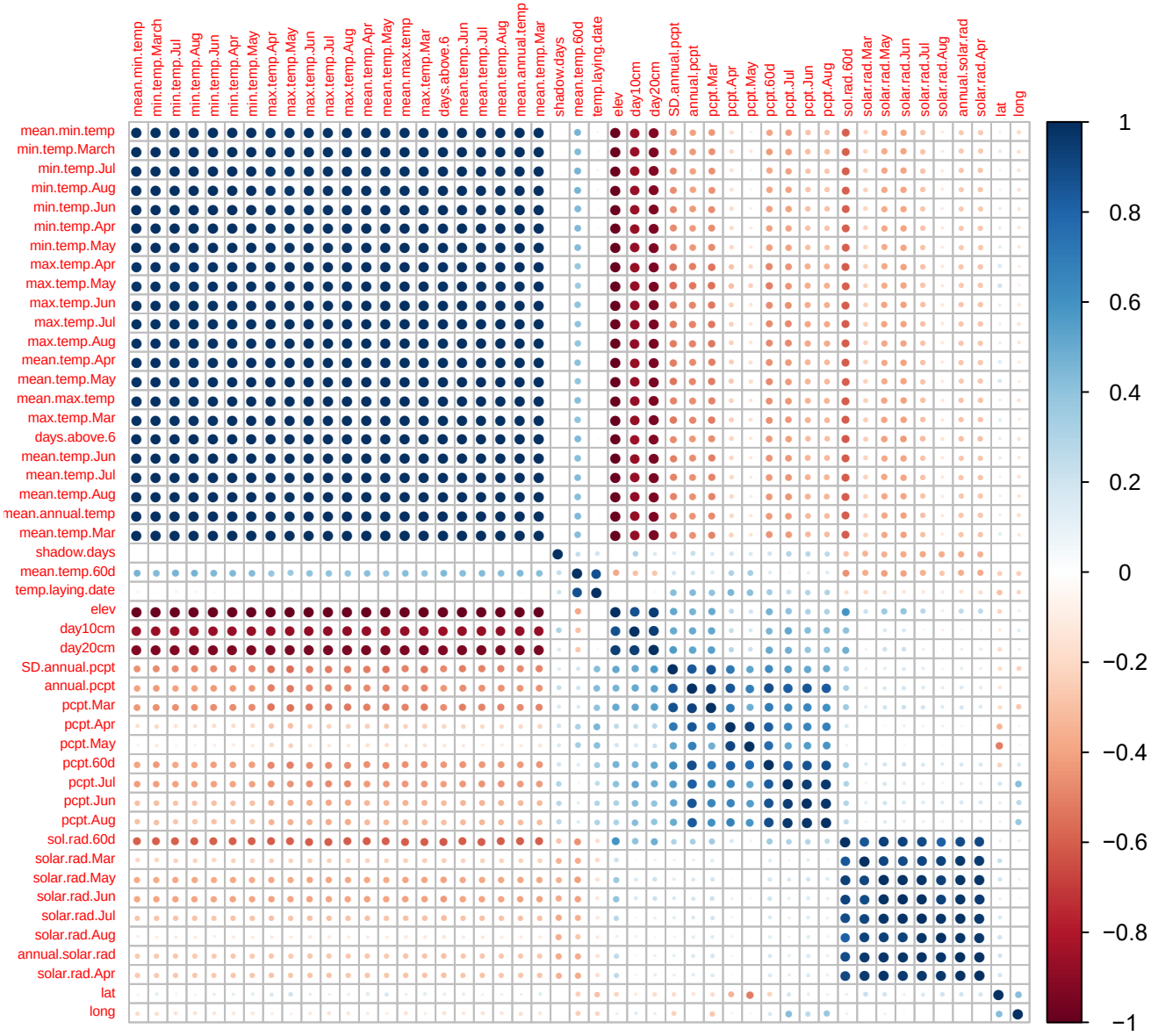


Figure S4.2 Correlation between all environmental variables obtained for sampled sites. Scale and colour as in Fig. S4.1.

Table S4.1 Partitioning of genetic variance in each geographic region using RDA and partial RDA analyses. The column "Partitioned variance" shows the total variance of the genetic data (Total variance), the proportion of variance that could be expained by the full RDA model which includes all spatial and climate explanatory variables (Full model: Spatial + climate), the proportion of total variance explained by the partial RDA when only climate variables are included and spatial variables are kept constant (Pure Climate), and the proportion of total variance explained by spatial variables when climate variables are kept constant (pure geog). The proportion of variance explained by spatially structured climate was calculated as the difference between the variance explained by the full model and that explained by the two partial RDA models. The column "Proportion constrained" shows the variation explained by each model relative to the total explainable variance. The fit (R^2adj) and significance (P) of each model is shown, and with significant p-values shown in bold.

CHall

Genetic Variation	Partitioned variance	Proportion constrained	R^2adj	P
Total variance	0.339			
Full model: Spatial + climate (constrained variance)	0.194	1.000	0.431	**0.001**
Pure Climate(climate|spatial)	0.017	0.086	0.017	**0.006**
pure geog (geog|climate)	0.165	0.848	0.405	**0.001**
Spatial + climate	0.013	0.066	0.010	NA

CH-North

Genetic Variation	Partitioned variance	Proportion constrained	R^2adj	P
Total variance	0.215			
Full model: Spatial + climate (constrained variance)	0.132	1.000	0.229	**0.001**
Pure Climate(climate|spatial)	0.054	0.406	0.045	0.129
pure geog (geog|spatial)	0.055	0.416	0.117	**0.004**
Spatial + climate	0.024	0.178	0.067	NA

CH-South

Genetic Variation	Partitioned variance	Proportion constrained	R^2adj	P
Total variance	0.308			
Full model: Spatial + climate (constrained variance)	0.157	1.000	0.248	**0.001**
Pure Climate(climate|spatial)	0.062	0.397	0.045	**0.057**
pure geog (geog|spatial)	0.052	0.328	0.088	**0.002**
Spatial + climate	0.043	0.275	0.115	NA

CH-East

Genetic Variation	Partitioned variance	Proportion constrained	R^2adj	P
Total variance	0.257			
Full model: Spatial + climate (constrained variance)	0.125	1.000	0.230	**0.001**
Pure Climate(climate|spatial)	0.029	0.233	0.008	0.261
pure geog (geog|spatial)	0.045	0.362	0.106	**0.001**
Spatial + climate	0.051	0.405	0.116	NA

CH-SouthWest

Genetic Variation	Partitioned variance	Proportion constrained	R^2adj	P
Total variance	0.216			
Full model: Spatial + climate (constrained variance)	0.185	1.000	0.351	**0.008**
Pure Climate(climate|spatial)	0.132	0.714	0.322	0.088
pure geog (geog|spatial)	0.058	0.311	0.275	0.187
Spatial + climate	-0.005	-0.026	-0.246	NA

CH-SouthEast				
Genetic Variation	**Partitioned variance**	**Proportion constrained**	**R^2adj**	**P**
Total variance	0.289			
Full model: Spatial + climate (constrained variance)	0.202	1.000	0.015	0.471
Pure Climate(climate|spatial)	0.089	0.441	-0.103	0.803
pure geog (geog|spatial)	0.080	0.397	-0.043	0.628
Spatial + climate	0.033	0.162	0.161	NA

Table S4.2 Summary of the contribution of significant predictor variables to explaining genomic variation. Significant variables were identified using a forward selection approach on the full RDA model (i.e. including all spatial and climate variables as predictor variables). The squared value of RDA constraint scores for climate and spatial variables for all sites (A), global (B-D), and fine-scale (E-F) geographic regions are shown. Grey hue increases with increasing association with that RDA axis. The proportion of total genomic variation explained by each axis is shown in parentheses.

A) CHall

Variable	RDA1 (30%)	RDA2 (7.8%)	RDA3 (3.6%)	RDA4 (2%)	RDA5 (1.4%)	RDA6 (1.2%)
PCNM1	0.00	0.92	0.01	0.01	0.01	0.00
PCNM2	0.14	0.01	0.50	0.20	0.00	0.02
PCNM3	0.05	0.02	0.01	0.33	0.22	0.01
PCNM4	0.39	0.00	0.03	0.02	0.05	0.00
PCNM5	0.03	0.01	0.14	0.10	0.28	0.13
PCNM6	0.01	0.01	0.07	0.00	0.02	0.00
PCNM8	0.20	0.01	0.00	0.25	0.15	0.14
PCNM9	0.07	0.01	0.09	0.09	0.14	0.13
PCNM10	0.04	0.01	0.14	0.00	0.04	0.00
sol.rad.60d	0.01	0.01	0.02	0.04	0.10	0.00
pcpt.60d	0.02	0.00	0.05	0.07	0.03	0.63

B) CH-North

Variable	RDA1 (13.5%)	RDA2 (10.3%)	RDA3 (8.3%)	RDA4 (5.5%)
PCNM1	0.00	0.81	0.06	0.13
PCNM3	0.26	0.00	0.47	0.27
PCNM4	0.00	0.19	0.27	0.54
day10cm	0.69	0.05	0.14	0.12

C) CH-South

Variable	RDA1 (20.9%)	RDA2 (6.0%)
PCNM1	0.98	0.02
PCNM2	0.02	0.98

F) CH-East

Variable	RDA1 (19.3%)	RDA2 (6.6%)	RDA3 (3.4%)	RDA4 (2.8%)	RDA5 (2.4%)
PCNM1	0.43	0.00	0.13	0.43	0.01
PCNM2	0.00	0.99	0.00	0.00	0.01
PCNM3	0.50	0.00	0.31	0.14	0.05
PCNM4	0.05	0.01	0.08	0.22	0.64
pcpt.60d	0.02	0.02	0.74	0.10	0.12

D) CH-SouthWest

Variable	RDA1 (22.3%)	RDA2 (14.1%)	RDA3 (12.1%)	RDA4 (10.6%)	RDA5 (7.6%)
PCNM1	0.13	0.10	0.57	0.20	0.00
shadow.days	0.16	0.81	0.00	0.02	0.00
sol.rad.60d	0.15	0.04	0.37	0.11	0.32
pcpt.60d	0.21	0.07	0.51	0.21	0.00
temp.laying.date	0.19	0.10	0.17	0.10	0.45

E) CH-SouthEast

Variable	RDA1 (12.9%)	RDA2 (6.7%)
sol.rad.60d	0.32	0.68
pcpt.60d	0.49	0.51

Table S4.3 Summary of the significant climate variables identified using forward selection on a partial RDA, where spatial variables were partialled out of the model. As in Table 2, the squared term of the constraint scores of each variable are shown. The proportion of total genomic variance explained by each RDA axis is shown in parentheses.

A) CHall

Variable	RDA1 (1.8%)	RDA2 (0.9%)
Solar radiation during 60 days after spawning	0.13	0.70
Mean precipitation during 60 days after spawning	0.12	0.46

B) CH-North

Variable	RDA1 (6.1%)
Mean precipitation during 60 days after spawning	0.75

C) CH-South

Variable	RDA1 (5.4%)	RDA2 (4.0%)	RDA3 (3.7%)
Solar radiation during 60 days after spawning	0.33	0.15	0.22
Mean precipitation during 60 days after spawning	0.00	0.23	0.00
Mean temperature on estimated spawning date	0.06	0.12	0.71

D) CH-East

Variable	RDA1 (3.1%)
Mean precipitation during 60 days after spawning	0.87

E) CH-SouthWest

Variable
~1

F) CH-SouthEast

Variable
~1

Table S4.4 Summary of the significant climate variables identified in each geographic region using forward selection on a partial RDA with climate variables were partialled out of the model. The squared term of the constraint scores of each variable are shown. The proportion of total genomic variance explained by each RDA axis is shown in parentheses.

A) CHall

Variable	RDA1 (27.6%)	RDA2 (6.8%)	RDA3 (2.9%)	RDA4 (1.4%)	RDA5 (1.2%)	RDA6 (0.8%)
PCNM1	0.00	0.91	0.01	0.00	0.00	0.01
PCNM2	0.14	0.01	0.24	0.12	0.07	0.01
PCNM3	0.06	0.00	0.06	0.33	0.09	0.17
PCNM4	0.34	0.00	0.05	0.00	0.02	0.32
PCNM5	0.02	0.00	0.28	0.06	0.22	0.05
PCNM8	0.17	0.00	0.03	0.02	0.36	0.00
PCNM9	0.09	0.02	0.10	0.20	0.02	0.36
PCNM10	0.05	0.01	0.12	0.00	0.04	0.00

B) CH-North

Variable	RDA1 (9.3%)	RDA2 (6.6%)
PCNM1	0.23	0.58
PCNM3	0.61	0.14

C) CH-South

Variable	RDA1 (8.1%)
PCNM1	0.23

D) CH-East

Variable	RDA1 (6.6%)	RDA2 (6.1%)	RDA3 (2.5%)
PCNM1	0.19	0.00	0.35
PCNM2	0.04	0.88	0.00
PCNM3	0.03	0.03	0.27

E) CH-SouthWest

Variable
~1

F) CH-SouthEast

Variable
~1

Adaptive genomic variation associated with environmental gradients along a latitudinal cline

Abstract

Rana temporaria occur across a large geographic and environmental gradient in Scandinavia. Several studies involving common garden experiments have established adaptive divergence across the gradient. The main objective of this study was to determine the extent of neutral and adaptive genetic divergence across the latitudinal gradient. Here I sequence genome-wide markers for 15 populations from six regions sampled from southern Sweden to Finland. Using a multivariate approach I find that 68 % of the genomic variation is associated with climate or geographically structured climate. Using outlier scans and environmental association analyses I identify a set of potentially adaptive loci and examine their change in allele frequency associated with different climatic variables. Using a gradient forest analysis I identify points along three of the climate variables where allele frequencies change more rapidly than expected from a linear association. I identify a large threshold effect associated with BIO5 (mean temperature during the warmest month) which is seen as a rapid change in southern Sweden. By comparing the change in neutral and adaptive allele frequencies across the whole gradient, I identify southern Sweden as a region with the largest divergence between the datasets. This suggests small changes in the climate may result in a mismatch between the adaptive genotypes and the environment in these populations. Overall this study shows that genomic analyses can provide a powerful complement to common garden experiments to improve our understanding of adative divergence across heterogeneous landscapes.

Introduction

The geographic distribution of genetic variation across a species range is an important determinant of population persistence under changing environmental conditions (Hoffmann & Sgrò 2011). To mitigate future biodiversity loss due to climate change it is important that we identify the most important environmental drivers of adaptive divergence and determine how genetic variation contributes to adaptation. Landscape genomics methods have been employed to identify potentially adaptive loci across many study systems (reviewed in Rellstab *et al.* 2015). However, adaptation across environmental gradients are often characterised by divergence in multiple phenotypic traits with polygenic genetic underpinnings (Pritchard & Di Rienzo 2010; Yeaman 2015). This presents a challenge when examining non-model organisms, where the lack of genomic resources is often prohibitive in determining the genomic architecture and genetic basis of adaptation (Manel *et al.* 2016). However, by developing our understanding of the distribution of genetic variation (adaptive and neutral) in geographic and climate space, we can draw meaningful conclusions about the ecological determinants of species distributions and adaptive divergence without pinpointing the underlying causal variants (Jones *et al.* 2013; Fitzpatrick & Keller 2015; Forester *et al.* 2016).

Two main statistical models have been developed to identify potentially adaptive loci while accounting for population structure (reviewed in Hoban *et al.* 2016). Population genetic methods identify loci with higher than expected differentiation (usually measured by F_{ST}) compared to neutral expectations based on population structure (Luikart *et al.* 2003). These methods are effective for detecting selective sweeps associated with strong selection acting on a few beneficial loci. However, the signal of divergence is difficult to detect if there is high gene flow between populations adapted to different conditions, or when the adaptive traits are polygenic with many small changes in allele frequency additively contributing to adaptation (Kawecki & Ebert 2004; Pritchard & Di Rienzo 2010). The second kind of model, termed Environmental Association Analysis (EAA; reviewed in Rellstab *et al.* 2015), is aimed at finding associations between allele frequencies and environmental variables, thus does not rely on strong sweeps underlying adaptation. This approach

is particularly useful for comparing the importance of different environmental variables and for mapping spatial changes in allele frequencies of adaptive loci. However, EAA assumes a linear relationship between allele frequencies and environmental gradients (Thomassen *et al.* 2010; Fitzpatrick *et al.* 2015), thus confining inferences to linear responses. But non-linear patterns of genetic variation along environmental gradients are probably common, judging from laboratory studies of organismal physiology (Angilletta's 2009 thermal adaptation book, for example), and could be important for identifying populations or geographic regions that are especially vulnerable to climate change. Thus, it is important to modify EAA such that non-linear relationships between allele frequencies and environmental gradients can be detected in nature. Here we combine a modelling approach to loci identified using FST outlier and EAA methods to identify such non-linear associations with environment in the European common frog, *Rana temporaria* across a latitudinal gradient.

Rana temporaria is widespread across Europe and occur throughout Scandinavia (Sillero *et al.* 2014). Populations occur across a wide range of habitats, which suggests adaptive divergence across the species range. Common garden experiments across 1600-km of the Scandinavian latitudinal gradient have confirmed this by establishing extensive latitudinal variation in larval and adult life history traits (Merila *et al.* 2000; Laugen *et al.* 2003b, 2005a; Palo *et al.* 2003a; Lindgren & Laurila 2005). The geographic scale of the gradient provides an interesting system to investigate the genomics of adaptation to environment, because there is little gene flow between populations adapted to different environments.

The main objective of this study is to determine the extent of neutral and adaptive genetic divergence across the latitudinal gradient. Specifically, we aim to 1) characterise the neutral genetic structure, 2) determine the proportion of the genome associated with environmentally driven adaptive divergence, and 3) identify environmental thresholds to adaptation by examining the non-linear response of adaptive loci to climate variables. We were particularly interested in determining whether there are non-linear relationships between adaptive allele frequencies and environmental variables. Such a response would suggest that there is a threshold along that particular environmental variable that requires a larger change in allele

frequency than would be expected based on the gradual change in that environmental variable. The results have important implications for identifying populations and geographic regions that would be particularly vulnerable to changing environments.

Methods

Sampling & DNA extraction - To determine intra-specific population structure and adaptive variation, 163 individuals were collected from six geographic regions across ~1500 km of the Scandinavian latitudinal gradient (Fig. 5.1; Table 5.1). Each region was represented by one to three populations (where a population is a pond), for a total of 15 populations in the final dataset. At each sampling site, we sampled approximately 10 eggs from 20-30 freshly –laid clutches (less than two days old). The eggs were transported to the laboratory at Uppsala University where they were raised in separate containers kept in climate room at 16 °C. Tadpoles were raised to Gosner stage 25 (Gosner 1960), whereafter they were euthanized with an overdose of MS222, preserved in 96% ethanol and stored at 4 °C until DNA extraction. Total DNA was extracted from one individual per clutch (henceforth family) using the Qiagen DNeay blood and tissue kit (Qiagen, CA, USA).

ddRAD sequencing, de novo assembly, and variant calling - To establish a genome-wide marker set, double digest restriction-site-associated DNA sequencing libraries were prepared with the restriction enzymes *EcoRI* and *MseI*, using a modification of the protocol by Peterson *et al.* (2012). We constructed 6 libraries comprising 48 samples each for single-end sequencing (125bp) on an Illumina HiSeq 2500 v4 at the Functional Genomics Center, University of Zurich. Individual samples were identified by unique 5-bp barcodes. Raw sequence reads were demultiplexed using the process_radtags package from Stacks (Catchen *et al.* 2011). Demultiplexing was based on a unique 9-bp sequence for each individual (5-bp unique barcode + 4-bp restriction enzyme recognition site), with 1 mismatch allowed.

Adapter and other Illumina-specific sequences were removed using Trimmomatic v0.33 (Bolger *et al.* 2014).

De novo assembly and variant calling was implemented using pyRAD (Eaton 2014) which first finds clusters within individuals based on a clustering threshold and minimum depth, and then clusters these loci between individuals (using the same clustering threshold), and identifies loci found in a user specified number of individuals. Clustering thresholds of 90-99% sequence similarity were tested; an optimum of 94% was chosen because it maximised nucleotide diversity and minimised the estimated number of paralogs in the dataset. A maximum of 4 sites with a Phred quality score <20 were allowed per sequence. Clusters were kept if they had >5x coverage per individual and were found in at least 4 individuals.

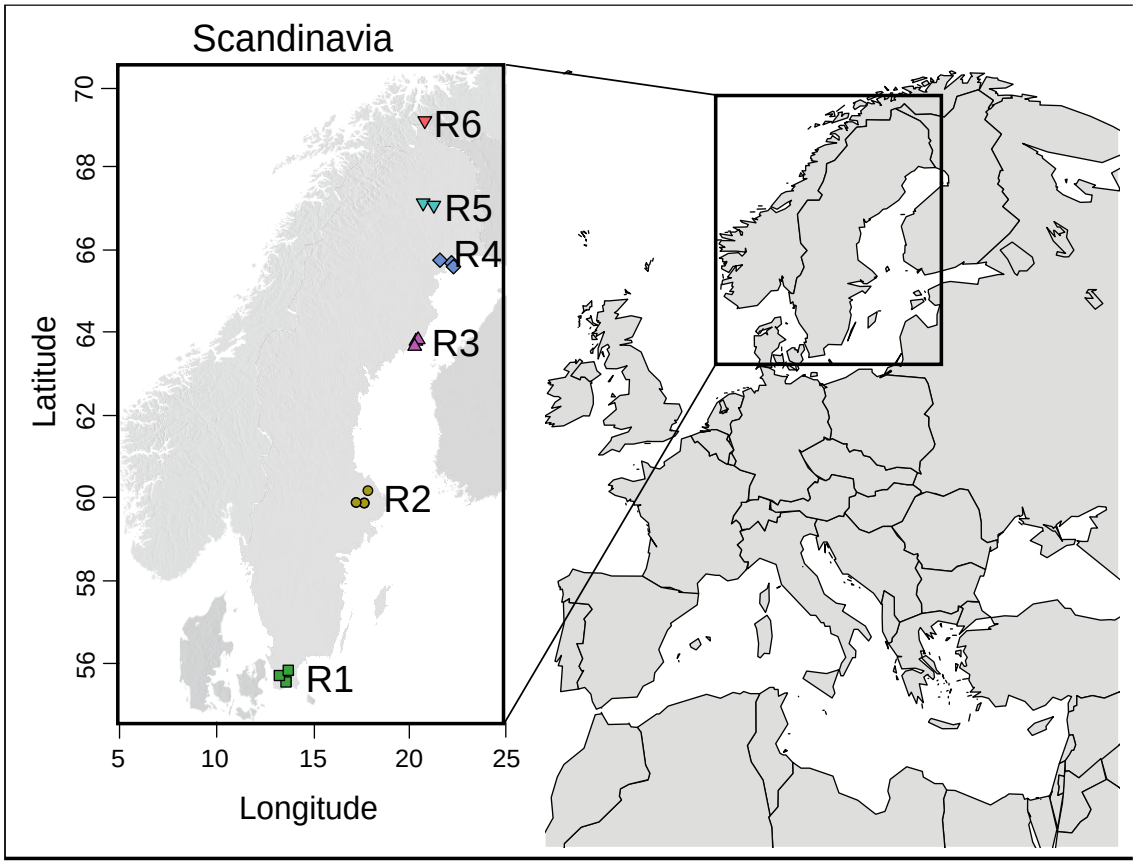


Figure 5.1 Sampled populations across the Fennoscandian latitudinal gradient. One to three populations were sampled within each of six regions. Symbols represent the five genetic clusters found at K=5, the most likely number of divisions found by DAPC, PCAdapt, and SNMF analyses. Geographic regions are named R1 - R6 from south to north along the latitudinal gradient.

Table 5.1 Summary of the diversity statistics calculated per population. The geographic coordinates (Lat, Long) are shown for each population (Pop). The geographic region (Region) to which the populations belong is shown along with the abbreviation (R1-R6) used throughout the manuscript. The number of individuals included in the analyses (n) are shown. Gene diversities (Hs), deviation from random random mating (F_{IS}), observed heterozygosity (Ho), and the average number of heterozygous sites across all sequenced sites (AvgHet) with standard deviation (SD) are reported per population.

Region	Pop	Lat	Long	n	F_{IS}	Hs	Ho	AvgHet	SD
All				132	0.10	0.23	0.21	0.0025	
	Sk.Ho	55.859	13.764	9	-0.01	0.29	0.29	0.0033	0.0004
R1: Skåne ■	Sk.SF	55.558	13.638	10	0.12	0.27	**0.23**	0.0030	0.0001
	Sk.SL	55.723	13.287	17	0.14	0.27	**0.23**	0.0031	0.0001
	Upp.Gra	59.878	17.667	9	0.07	0.19	**0.18**	0.0023	0.0001
R2: Uppsala ●	Upp.K	59.891	17.242	10	0.08	0.22	**0.20**	0.0024	0.0001
	Upp.O	60.178	17.854	9	0.07	0.21	**0.19**	0.0024	0.0001
	Um.Gr	63.792	20.367	9	0.07	0.21	0.19	0.0023	0.0001
R3: Umeå ▲	Um.Taf	63.830	20.486	10	0.09	0.22	0.20	0.0024	0.0001
	Um.UT3	63.658	20.298	2	0.01	0.20	0.18	0.0021	0.0003
	LT1	65.684	22.213	9	0.06	0.21	0.20	0.0024	0.0001
R4: Luleå ◆	LT2	65.750	21.602	3	0.03	0.24	0.21	0.0024	0.0001
	LT3	65.583	22.319	7	0.05	0.22	0.20	0.0024	0.0001
R5: Kiruna ▼	Kir.G	67.111	20.656	10	0.09	0.23	0.20	0.0025	0.0001
	Kir.L	67.052	21.224	10	0.08	0.22	0.20	0.0025	0.0001
R6: Finland ▼	FIN	69.044	20.805	8	0.10	0.22	0.19	0.0023	0.0002

SNP validation - The putative variants identified though the pyRAD pipeline were filtered for possible sequencing errors, paralogs, and uninformative SNPs. The following filters were applied: 1) SNPs that were genotyped in less than 50% of individuals were removed using the *--max-missing* function in VCFtools v0.1.12b (Danecek *et al.* 2011). 2) Loci with a minor allele frequency less than 0.05 in the full dataset were removed as they are more likely to be sequencing error, and if they are true variants they are uninformative and likely to bias tests for selection (Roesti *et al.* 2012). 3) Using a sliding-window of 10-bp we tested whether the number of variants increased towards the end of the sequence. No significant difference was found between bins, thus the sequences were not trimmed further. 4) We reduced linkage in the dataset we included only one variant per locus using the *--thin* function in VCFtools. 5) We assessed whether loci were in Hardy-Weinberg equilibrium (HWE) within each population using the *--hardy* function in *PLINK* (Purcell *et al.* 2007). Loci with an observed heterozygosity more than 0.5, and loci that deviated significantly from HWE based on the exact test ($p<0.05$; (Wigginton *et al.* 2005) were removed from the dataset if they occurred in more than 5 populations. 6) We then calculated linkage disequilibrium for each locus pair per population in PLINK v 1.07 (Purcell *et al.* 2007). Loci with $r^2 > 0.8$ never occurred in more than 5 populations, so no loci were excluded at this step. 7) Finally we excluded individuals with more than 55% missingness.

Summary statistics - Nucleotide diversity was calculated for each sample as the frequency of heterozygous bases (IUPAC codes) from the *pyRAD* output, and means were calculated per population. Although these calculations are based on sequences before final filters are implemented, two reasons convince us that these results are robust: 1) *pyRAD* calculates a binomial probability that a base is homozygous or heterozygous based on a maximum likelihood approach of jointly estimating the heterozygosity and sequencing error rate from the base frequencies within each individual's stacks. If the read depth of the stack falls below a user-set threshold, or is too low to make a statistical base call, the base remains undetermined. Thus the final base calls per individual should be fairly robust (Eaton 2014). 2) The post-*pyRAD* filters outlined in the previous paragraph do not address sequencing or SNP-calling

errors, but rather minimises missingness, systematic biases, and linkage between loci.

Further summary statistics were calculated in R v3.3.1 (R core team 2016) with the *hierfstat* v*0.04-22* and *adegenet v2.0.1* packages (Goudet 2005; Jombart 2008; Jombart & Ahmed 2011). Gene diversities (Hs), deviation from random mating (F_{IS}), and observed and expected heterozygosity are reported per population (Table 5.1).

Population structure - Pairwise population differentiation was estimated using Weir and Cockerham's F_{ST} (Weir & Cockerham 1984) as implemented in *hierfstat* (Goudet 2005). We visualised the genetic distance between populations with a principal component analysis (PCA) implemented in *PCAdapt v3.0.3* (Luu *et al.* 2016). The following analyses were conducted in *adegenet.* To test for isolation by distance, we calculated the correlation between log transformed pairwise geographic distances and scaled pairwise genetic distances (F_{ST}/1-F_{ST}) (Rousset 1997). We tested for significance using a Mantel test. We quantified the proportion of genetic variation that explained differentiation within and between populations with an analysis of molecular variance (AMOVA). Finally, we performed a discriminant analysis of principal components (DAPC) to determine the most likely number of clusters in the dataset and visualise broad-scale population structure.

Climate Data - We obtained climate data from WorldClim v1.4 (Hijmans *et al.* 2005) at a resolution of 2.5 minutes of degrees using the R package *raster v.2.5-8* (Hijmans *et al.* 2015). Many of these variables are derived from the same data and are highly correlated. To reduce the redundancy in the climate variables retained for analyses, we first calculated correlation between all variables, and then removed variables if they exceeded a correlation threshold. We calculated the absolute values of pairwise ranked correlation (Spearman's rho) between all 19 BioClim variables from WorldClim, longitude, latitude, and season length. Season length was calculated as the number of days above 6 and 8 °C at each sampling site, since 6 °C is approximately the development threshold of *R. temporaria* tadpoles (Laugen *et al.* 2003a). We reduced redundancy in the environmental dataset by detecting pairs of

variables that had an absolute correlation >0.8, and then eliminating the one that had the highest mean correlation with all other variables (Kuhn et al. 2016). This procedure retained five BioClim variables, and these were used as the environmental variables in all remaining analyses (Fig. 5.2).

Relative contribution of environment and IBD to genomic differentiation - We investigated the effects of climate and geography on neutral genetic structure using full and partial redundancy analyses (RDA) with variance partitioning. Redundancy analysis is a form of multivariate regression, which can be used when both the predictor and response variables are multivariate (Legendre & Legendre 2012). As a canonical extension of multiple linear regression, RDA identifies a set of orthogonal linear predictor variables that explains the most variation in a set of linear response variables. In this case each RDA axis represents a set of co-varying loci (response variables), which are correlated with co-varying environmental variables (predictor variables). RDA has greater power to detect multivariate genotype-environment relationships than methods based on distance matrices or Mantel tests (Legendre & Fortin 2010).

We created two matrices as response variables: 1) the 5 climate variables identified above, centered and standardised, and 2) geographic coordinates of each sampling site. The response matrix was the minor allele frequencies of 2081 loci for each individual. We ran a sequence of nested models to partition variation in climate and geography as explanatory variables of allele frequencies: (1) the full model including both climate and geography; (2) climate only, with the influence of geography partialled out (climate | geography); and (3) geography only, with climate partialled out (geography | climate). The difference in the variance explained by model (1) minus the sum of models (2) and (3) was interpreted as the contribution of climate and geography acting together. Overall and residual variance was calculated for each model, and the model significance was tested with 999 permutations. The RDA was conducted using the R package *vegan v2.4-1* (Oksanen *et al.* 2015).

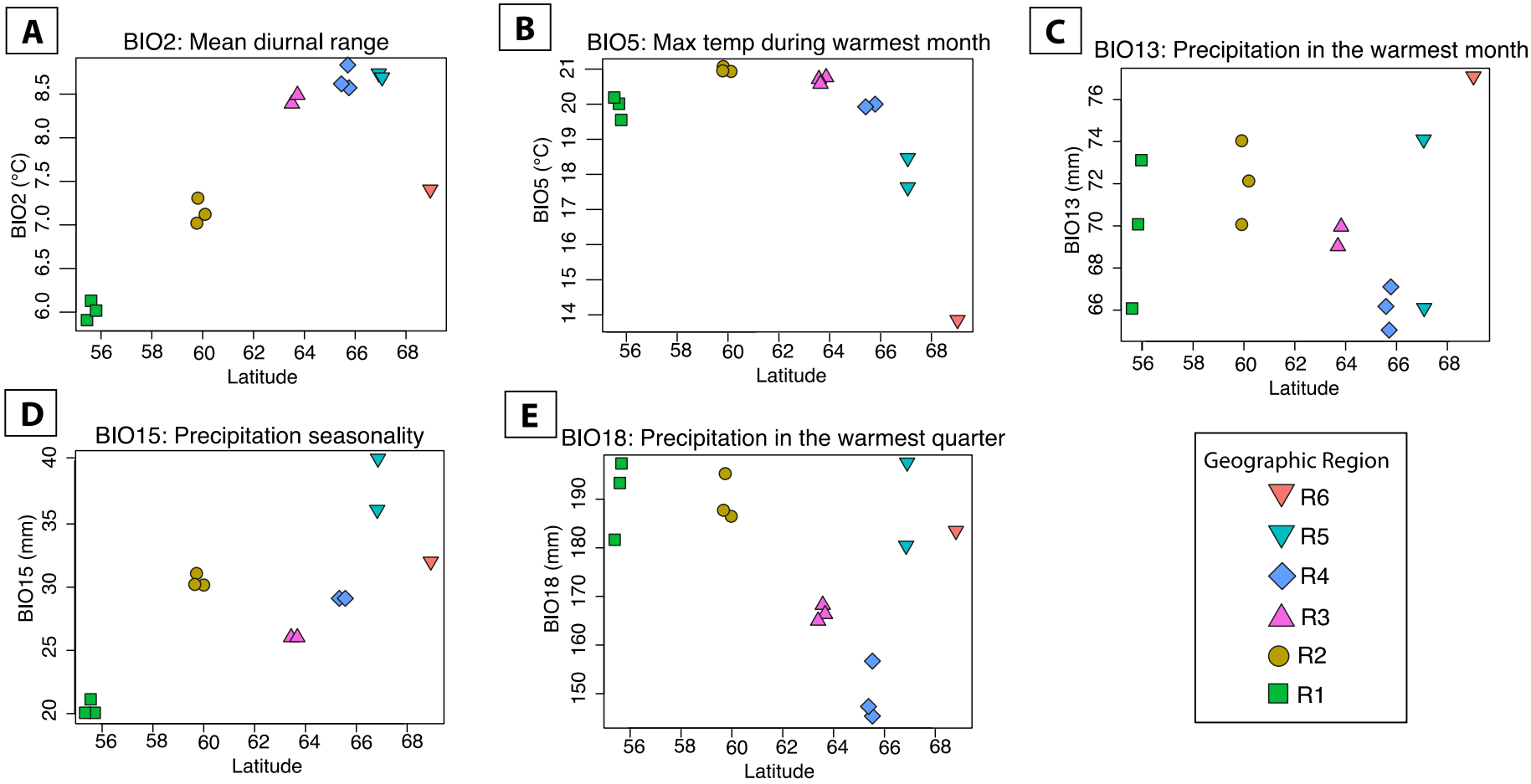


Figure 5.2 Latitudinal distribution of the five BioClim variables used in this study. Each point represents a sampled population.

Signatures of adaptation: Outlier detection and genotype-environment associations - We created two datasets containing loci potentially under selection using two common approaches: 1) Outlier analyses to identify loci that are more differentiated between populations than expected under a neutral model, and 2) Environmental Association Analyses (EAA) to identify loci strongly associated with an environmental variable (reviewed in Rellstab *et al.* 2015). These approaches are likely to identify different loci, since their underlying assumptions are different. The Outlier dataset comprised loci identified with *PCAdapt* (Luu *et al.* 2016), and from the X_TX statistic calculated in *bayenv2* (Günther & Coop 2013). The EAA dataset comprised loci identified using *bayenv2* and LFMM (Frichot *et al.* 2013).

PCAdapt identifies outlier loci as those that are more associated with population structure than expected. We used the R package *pcadapt 3.0.4* (Luu *et al.* 2016), which calculates a vector of z-scores of the how related each SNP is to the first K principal components, where K is the user-specified number of population clusters. A Mahalanobis distance is then calculated for each SNP to determine whether it deviates from the main distribution of z-scores. These scores are scaled by a constant, the genomic inflation factor, which produces a chi-squared distribution of values with K degrees of freedom. K was calculated as the most likely number of genetic clusters after testing K 1-20 and inspecting the scree plot of the proportion of explained variance for each K (Fig. S5.1). Based on these results, we chose K=5 for further analyses. We used a false discovery rate of 10% to identify outlier loci.

Bayenv2 estimates genotype-environment associations and an F_{ST}-like statistic (X_TX) while correcting for covariance of allele frequency between populations due to neutral processes. We used *bayenv2* to identify loci for both datasets. First we estimated the neutral covariance matrix based on 500 randomly selected loci. Two independent runs with 100000 MCMC iterations were run. We tested for convergence within each run by calculating Pearson's product-moment correlation (*cor.test* in R) between the final matrix and nine matrices printed out at 10000 step intervals (9 correlations). We constructed distance-based trees to determine whether relationships among populations remained constant within and between runs. Convergence between runs was calculated as the correlation between the final matrix in each run. Since these were highly correlated, we arbitrarily chose the final

matrix from the first run as our final covariance matrix. The full model was run using this covariance matrix, a file containing standardised measures of each environmental variable, and a genotype file containing SNP counts across all populations. We ran a non-parametric test that calculates the Bayes factor, Spearman's *p,* and Pearson's correlation coefficient for each genotype-environment association (-t -c -r). In addition we calculated the X_TX population differentiation statistic (-X). For this test, Gunther *et al.* (2013) suggest ranking loci by their X_TX statistic rather than selecting those above a specific threshold.

We conducted three independent runs with *bayenv2* of 100,000 MCMC iterations each for each of the five genotype-environment associations, and tested convergence by calculating the correlation between runs for each statistic (BF, *p*, X_TX). We also compared the overlap in loci identified in the top 5%, 6-10%, and 11-15% ranked loci based on the X_TX statistic for each environmental variable to ensure the repeatability of the results. We then ran an additional 7 independent *bayenv2* runs, and calculated the median result across all 10 runs as our final output. We selected the top 100 ranked loci based on the X_TX statistic for the outlier dataset. For the EAA dataset we selected loci with a log10 Bayes Factor (BF) >0.5 (Kass & Raftery 1995), and absolute Spearman's rho (*p*) >0.3.

Finally, we screened the dataset for additional EAA loci using the Latent Factor Mixed-effect Model (LFMM; Frichot *et al.* 2013; Frichot & François 2015). LFMM calculates the correlation between genotype and environment while simultaneously accounting for population structure with latent factors incorporated in the model. The number of latent factors is user specified, and should represent the number of genetic clusters (K) that best describes the population structure in the dataset. As suggested by Frichot & François (2015), we estimated the most likely K by evaluating the cross-entropy criterion for K1-10 using the function *snmf* in the R package *LEA*. The most likely K was 5, which is consistent with the population structure analyses described above, and therefore LFMM was run with K=5 for each of the five environmental variables. The Gibbs sampler algorithm was run five times for each environmental variable, with 10000 cycles and a burn-in of 5000 cycles. The median of the resulting correlation scores (z-scores) was calculated across all five runs. The authors suggest a recalibration of the mean z-scores by lambda; that is the

square of the mean z-scores divided by the median of a chi-squared distribution with one degree of freedom (λ; ~0.455). Lambda should be close to one − but more importantly, this should produce the correct adjusted p-value frequency distribution. Adjustment of λ can correct for liberal or conservative p-value distributions. We evaluated the effect of λ = 0.45-1.00 on the p-value distribution for each of the 5 environment-genotype associations. The shape of the distribution did not change much, but the frequency of p-values >0.1 increased as λ increased (i.e. the correction was more conservative). Thus, for a lambda close to one and the correct adjusted p-value distribution, we chose λ = 0.85 for BIO2, BIO5, BIO15, and BIO18, and λ = 0.45 for BIO13. To control for false discoveries, we applied a Benjamin-Hochberg adjustment with a false discovery rate of 5%.

Genomic Turnover across ecological gradients - We assessed how genomic variation changes across Scandinavia and whether important climatic thresholds occur by fitting a Gradient Forest model (Ellis *et al.* 2012) to each of the SNP datasets. We compare the change in allele frequency between the adaptive loci (Outlier and EAA dataset) and a Reference dataset composed of all the remaining loci. The Gradient forest model was developed as an extension of the random forest model to assess community level responses to ecological gradients, and has recently been applied to genomic data to detect non-linear change in allele frequencies along ecological gradients (Fitzpatrick & Keller 2015). It is a machine-learning ensemble approach that fits multiple regression trees between allele frequency and environmental variables. A set of decision trees is built to describe change in allele frequency across the predictor variable range. Each split is determined by minimising the "impurity" in the data, i.e. minimising the sums of squares of the allele frequency and thus maximising the tree fit. This means the split will always describe the biggest change in allele frequency at the current point in the tree, and a relative split importance can be calculated.

The response variable was the minor allele frequencies (MAF) for each SNP dataset. We included only loci variable in more than 4 populations to ensure robust regressions. The predictor variables were the 5 BioClim variables described above. To account for unsampled geographic structure in the dataset, we also included

Moran's Eigenvector Map variables (MEM; Dray et al. 2006), which are orthogonal vectors maximising spatial autocorrelation between sampled locations. Broad-scale spatial structure is most likely explained by the most positive eigenvectors (Manel et al. 2010b, 2012; Sork et al. 2013), so we included the first half of the positive MEMs here; in total three MEM vectors.

The gradient forest model was fit to each SNP dataset using the R package *gradientForest* (Ellis *et al.* 2012). We constructed 2000 regression trees per SNP, with default values for the variable correlation threshold (0.5), the number of candidate predictor variables sampled at each split (2), and the proportion of samples used for training (~66%) and testing (~33%) each tree. The relative importance (R^2) of each predictor variable was calculated as the weighted mean of the proportion of variance explained by the validation data. The cumulative importance for the change in allele frequency for each locus was calculated as the sum of the split importance across climatic variables, and the mean allelic turnover per climatic variable was calculated for each of the three SNP datasets.

Changes in allele frequency across the landscape were visualised by transforming each climatic variable by the genomic importance calculated for each SNP dataset; i.e. we produced a transformed dataset for the Reference, EAA, and Outlier datasets. The three transformed datasets were produced for climate data extracted from a raster stack covering eastern Sweden and northernmost Finland. For each dataset, the transformed variables were reduced by PCA, and a colour from the RGB colour palette in R was assigned to each of the first three principal components. Thus, for each geographic point along the latitudinal gradient, a single colour was used to represent the genomic composition in three-dimensional principal component space. To compare the genomic turnover between the neutral and two candidate SNP datasets, we calculated the distance in genomic space at each geographic point as the Procrustes residuals between the pairs of transformed matrices calculated above. The genomic difference between datasets was normalised and mapped in geographic space as before.

Results

ddRAD data generation and variant filtering - The final dataset consisted of 132 individuals from 15 populations (2-17 individuals per population; Table 5.1). There were 2081 SNP loci, with a mean depth of 17.9 - 25.5x and genotyping rate of 72-97% per population. A summary of the number of raw reads, the output from pyRAD, and the final dataset can be found in Table S5.1.

Genetic variation and Population structure - The highest genetic diversity was found in the southernmost region, R1, while the rest of the gradient was characterised by lower genetic diversity that was similar across all regions (Table 5.1). Specifically, the mean gene diversity (H_S) and heterozygosity (H_E) in Skåne were 0.28 and 0.25, respectively, while they were only 0.22 and 0.19 across the rest of the gradient. Similarly, the frequency of heterozygous sites across all sequenced sites averaged ~1/300 (0.0031) in R1, and ~1/400 (0.0024) across the rest of the gradient.

Measures of genetic differentiation showed evidence of population structure within and between regions. AMOVA indicated that most genetic variation (68.72%) was found within populations, while significant variation was found among populations within regions (6.05%) and among (25.23%) regions (Table S5.2). The mean global F_{ST} was 0.21 (Fig. 5.3, Table S5.3), suggesting strong population structure on average between populations. However, pairwise genetic distance was much lower within (F_{ST} = 0.06) than between (F_{ST} = 0.16) regions, and there was significant isolation by distance (R=0.434, p=0.001) across the sampled area.

To determine the broad-scale population structure, we first visualised the genetic distance between populations using PCA, and then estimated the most likely number of genetic clusters using DAPC. The first two axes of the PCA explained approximately 24% of the variance. PC1 (~15%) partitioned the two southern regions (R1 and R2) from the rest, and PC2 (~9%) partitioned populations latitudinally, with a graded differentiation from R2 to R6 (Fig. 5.4).

Discriminant analysis of principal components (DAPC) predicted five genetic clusters, corresponding to R1, R2, R3, R4, and the three northern populations (Fig. 5.1). When separating the dataset into increasing numbers of clusters, populations grouped out sequentially from south to north, except that R5 and R6 always grouped together.

Redundancy Analysis - The full RDA model included climate and geographic coordinates and explained 76.6% of the total genetic variation (*p=0.001*). Based on the partial RDA, climate was significantly associated with genetic variation (climate | geography; *p=0.002*), and explained 49.5% of the total variation. The variation explained by geography alone (geography | climate) was much less (11.4%) and was non-significant. The proportion of genetic variation explained by spatially structured climate (climate ∩ geography) was 39.1%.

Signature of adaptation: Outlier detection and genotype-environment associations - *PCAdapt* identified 50 outlier SNP loci and the X_TX statistic from *bayenv2* returned the top 100 outlier loci. A total of 28 loci were identified by both methods, so that the Outlier dataset comprised 122 unique loci (Fig. S5.2).

The EAA dataset comprised loci identified using *Bayenv2* and LFMM as associated with the five chosen BioClim variables. *Bayenv2* identified 123 unique loci (6% of the total loci tested), with 13% of these loci associated with multiple environmental variables (Fig. S5.3). LFMM identified 398 unique loci (~19% of the total loci), with ~30% associated with multiple environmental variables (Fig. S5.4). Only 22 loci were identified by both LFMM and *Bayenv2*; thus, the final EAA dataset comprised 499 unique loci (Fig. S5.5).

There were 56 loci present in both the Outlier and EAA datasets (Fig. S5.6). This represents 45.9% of the Outlier dataset, and 11.2% of the EAA dataset.

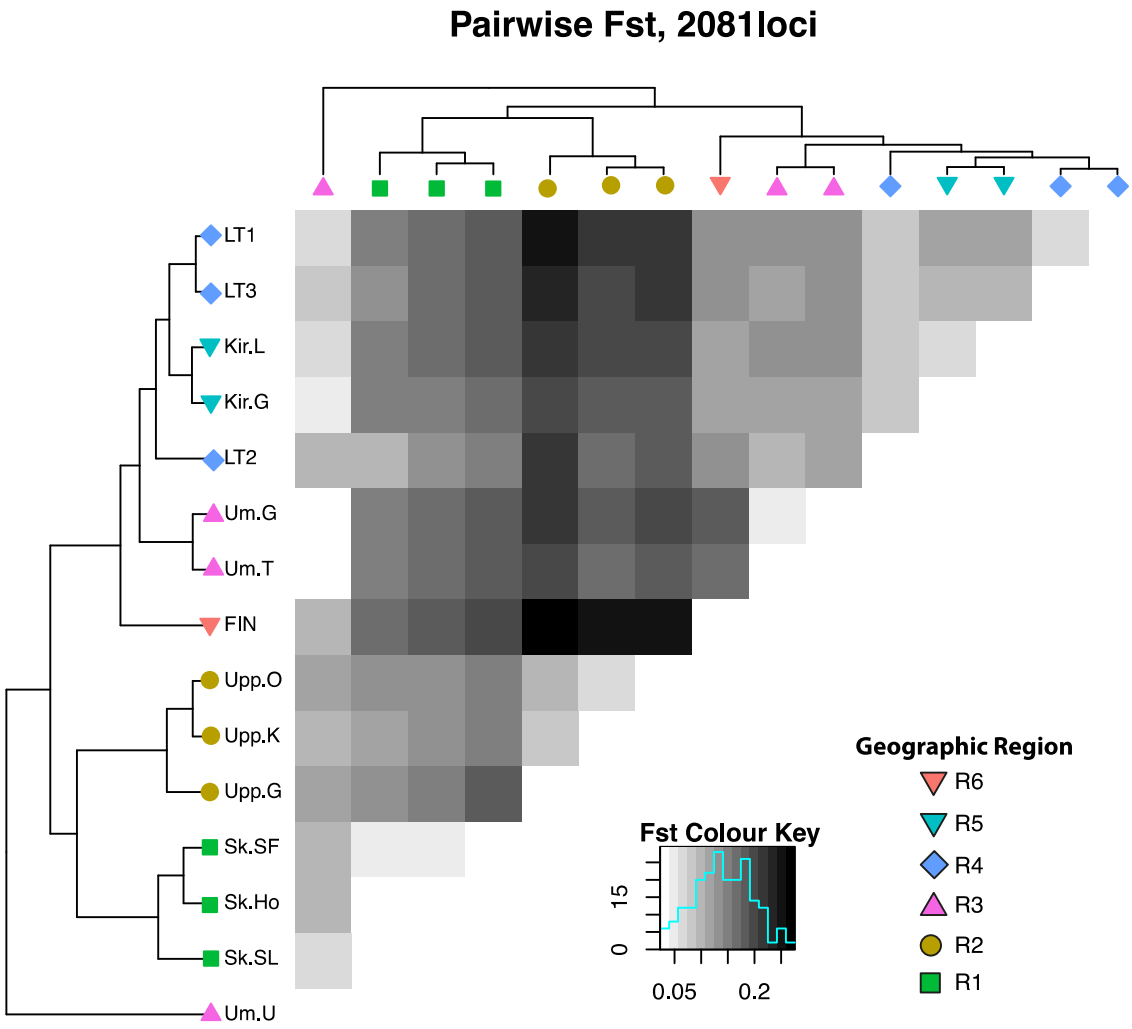


Figure 5.3 Heatmap of the pairwise genetic divergence (F_{ST}) between all sampled populations. Darker squares indicates higher F_{ST}, with colour scaled as shown in the key. Geographic regions are differentiated with shapes corresponding to populations in Fig. 5.1. Colours correspond to the PCA shown in Fig. 5.4. The dendrogram shows the population structure between southern (R1-R2) and northern (R3-R6) Scandinavian populations.

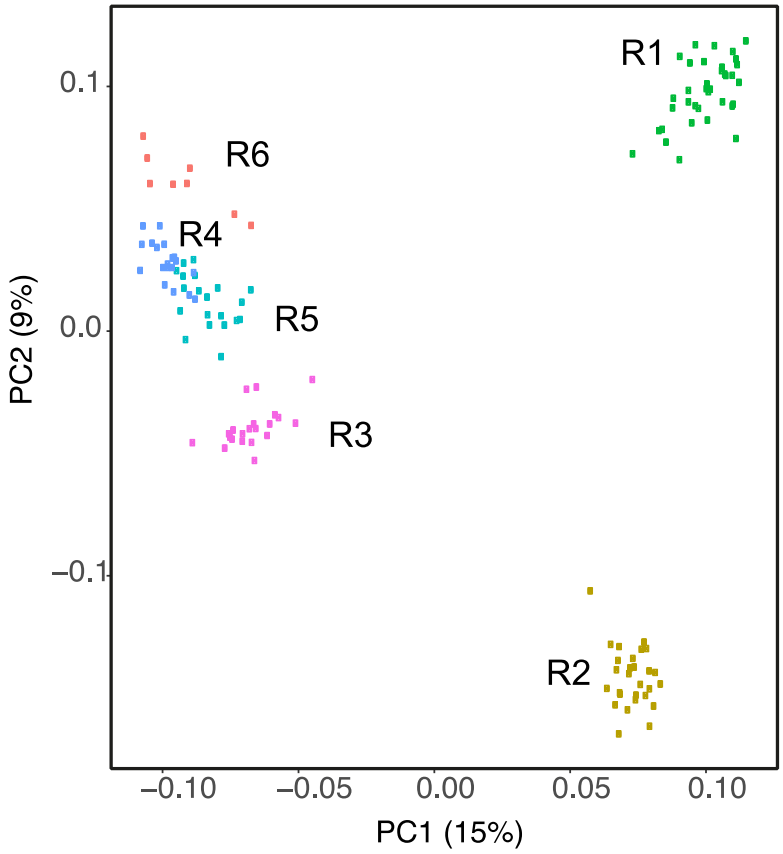


Figure 5.4 Graph of the first two axes of a Principal Component Analysis of all populations. The proportion variation explained by each axis is shown in brackets. Regions are shown in different colours.

Genomic Turnover - We used the mean R^2 averaged over loci in each dataset as a measure of the fit importance and thus how informative each dataset was (Table 5.2). The Reference and EAA datasets performed similarly, with mean R^2 values of 37.6% and 38.9%, respectively. The Outlier dataset showed a better fit on average, with a mean R^2 of 56.5%. The frequency distributions of R^2 values differed among datasets. The Reference and EAA datasets both had fairly flat distributions, with 30.5% and 34% of loci with $R^2>0.5$. The R^2 of the Outlier dataset was right skewed, with 67% of loci identified with $R^2>0.5$.

Table 5.2 Summary of the three SNP datasets and their results from the Gradient Forest analysis. The model was fit only to SNPs that were variable in more than five populations.

SNPs	nr of SNPs	polymorphic in >5 populations	SNPs with R^2 >0.5 (%)	mean % (range)
Reference	1427	1034	292 (30.5%)	37.6 (0.2, 90.0)
Outlier	122	115	77 (67.0%)	56.5 (0.8,91.5)
EAA	499	339	115 (33.9%)	38.9 (0.0,79.0)

SNPs, Single Nucleotide polymorphism

The most important variables for all three datasets were distance and either MEM1 or MEM2 (Fig. 5.5). This suggests that geographic distance is an important determinant of the genomic differences between populations, and that the MEM variables captured important environmental variation that had not been included in the model. When considering only the BioClim variables, BIO5 and BIO2 were the most important variables for the Outlier dataset, and BIO5 for the EAA dataset. Notably, the Outlier dataset had a higher fit importance (R^2) to the BIO2 and BIO5 variables than the EAA or Reference dataset. This suggests that these variables explain the change in allele frequency in the Outlier dataset better than the change in allele frequency in the Reference our EAA datasets.

We found that three BioClim variables explained a more rapid change in allele frequencies in the adaptive SNP datasets than in the reference datasets. The cumulative importance of allele frequency changes in the Outlier and EAA datasets differed in shape and magnitude for BIO5 (maximum temperature during the warmest month), BIO18 (precipitation during the warmest quarter), and MEM2 (Fig. 5.6 panels A, D, H). Most notably, the cumulative importance of the Outlier dataset showed big changes in allele frequencies at two points (19°C and 20.5°C) along BIO5 (maximum temperature during the warmest month). A similar, but slightly weaker response was seen in the EAA dataset (arrows in Fig. 5.6A). To examine this more closely we plotted the change in minor allele frequencies of five of the Outlier loci with the highest relative importance (R^2) associated with BIO5 (Fig. 5.7A). We find that allele frequencies in R2 and R3 differ dramatically from the rest of the gradient, while

frequencies in R1 are similar to R4-R6. This suggests a threshold response to the the higher temperatures experienced by R2 and R3 populations (Fig. 5.7B), where the adaptive genotype requires a big change in allele frequencies compared to the rest of the gradient. The rate of genomic turnover was less dramatic in response to BIO18 (precipitation in the warmest quarter, mm), but Outlier and EAA datasets both showed a higher cumulative importance than the Reference data (arrow in Fig. 5.6D). Finally, the Outlier dataset showed a higher cumulative importance associated with BIO2 (mean diurnal range in temperature) that was not recovered by the EAA dataset.

Spatial mapping of the model for all three datasets showed that genomic turnover was maximal between R1 and R2, followed by the change between R2 and the rest of the populations. The three precipitation variables explained most of the variation in the Reference data allele frequencies (biplots in Fig. 5.8A, C, and E), while BIO5 (maximum temperature during the warmest month) was important for both the EAA and Outlier datasets. The biggest difference in genomic turnover between the adaptive loci and the Reference dataset was found in southern Sweden (Fig. 5.9), which suggests that the biggest change in adaptive allele frequencies is required between populations from R1 to R2.

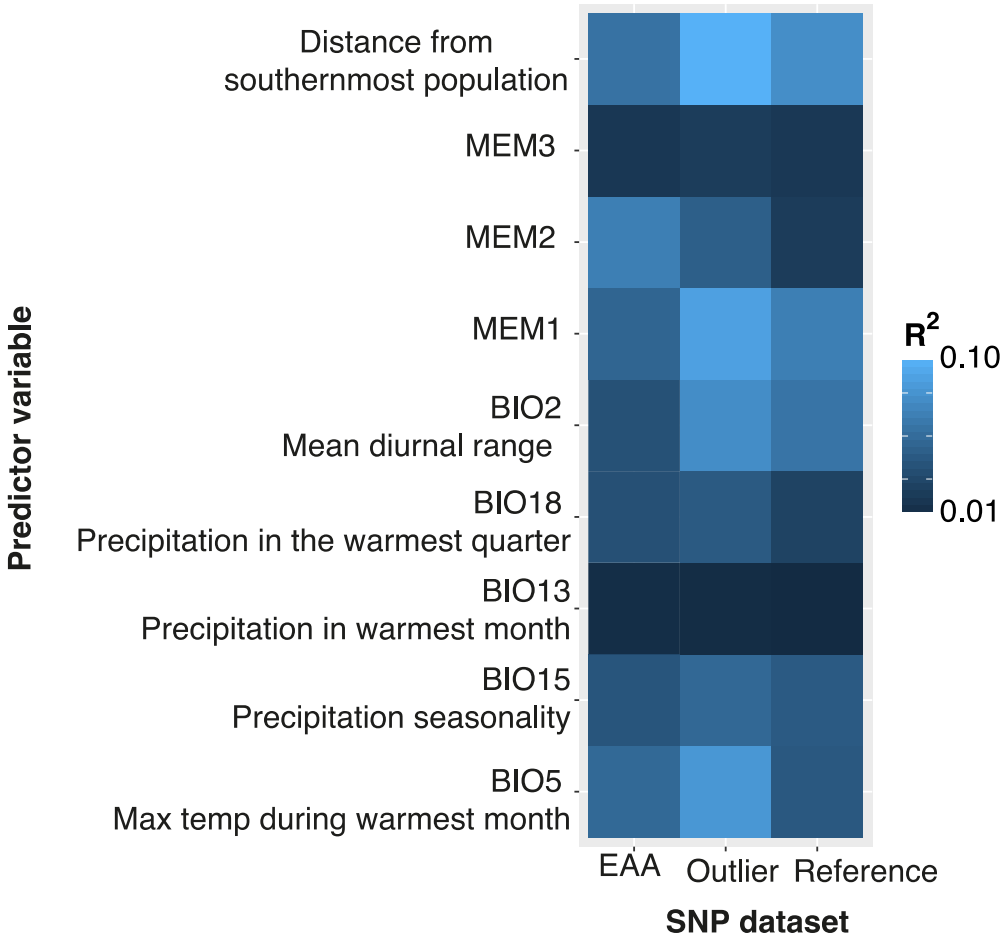


Figure 5.5 The relative importance (R^2) of the predictor variables used in the Gradient Forest analysis, calculated as the weighted mean proportion of variance in allele frequency explained by a given environmental variable. Results are shown for the five BioClim variables, geographic distance, and three Moran's Eigenvectors (MEM1-3) explaining geographic structure in the dataset. The columns show results for the SNP datasets. EAA: Adaptive loci from environmental association analyses; Outlier: Adaptive loci identified using PCAdapt and X_TX statistic in BayEnv2; Reference: the remaining loci. Lighter colours indicating higher importance; e.g. BIO2 and BIO5 explain a large proportion of variance in the Outlier dataset. The relative importance of variables associated with the Reference dataset provides the null model. Here BIO5 and BIO2 explain more variance in the Outlier SNP dataset than in the Reference dataset, which suggests that these environmental variables are important drivers of adaptive divergence.

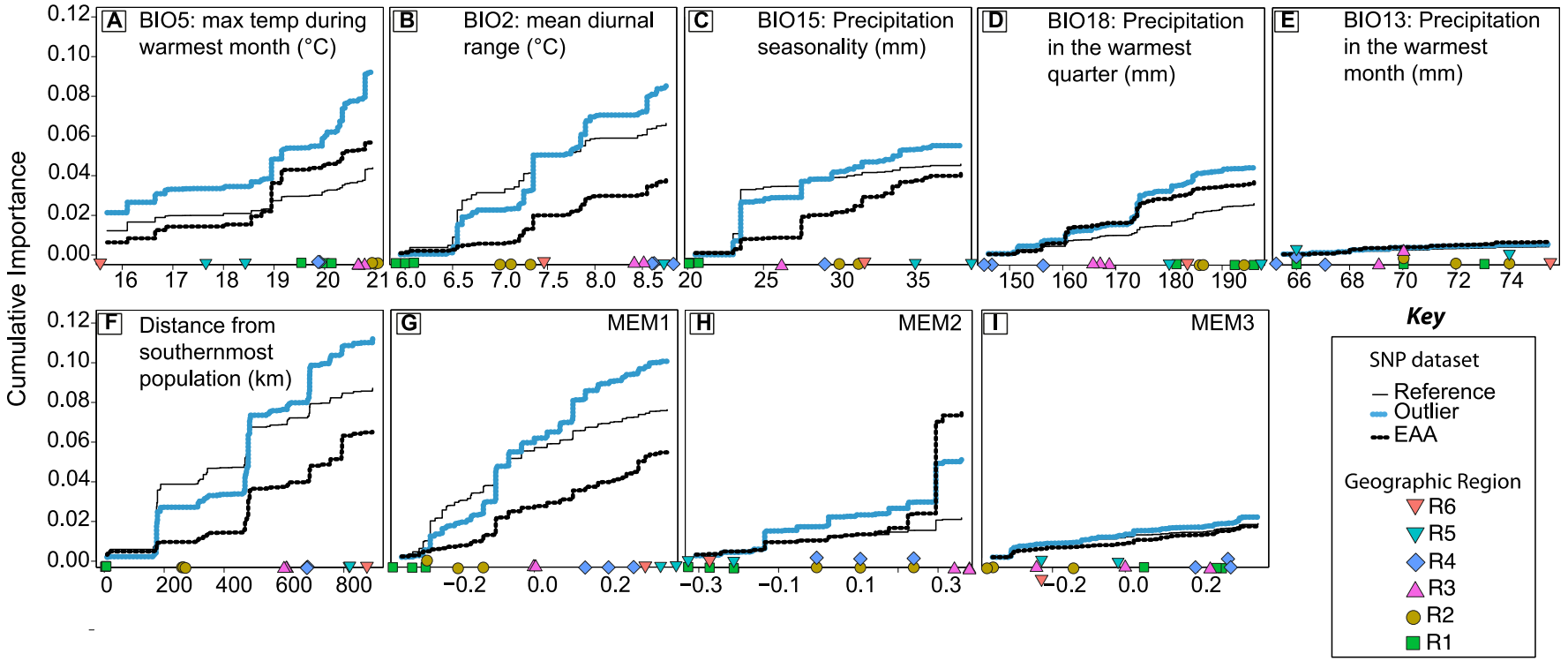


Figure 5.6 A comparison of the cumulative importance of each predictor variable for the three SNP datasets (see Key). We include the five BioClim variables (A-E) along with geographic distance (F) and three Moran Eigenvector Map variables (G-I) that explain geographic structure in the data. The maximum height of a line indicates the total allelic turnover associated with that variable. The relative importance of a particular point along the predictor variable is seen by the change in line height. The position of all populations along each variable is shown with coloured symbols along the bottom of each graph, with the populations from the same region following the same colour and shape codes as before (see Key). The Reference dataset shows the cumulative importance of each variable in explaining the change in allele frequency of neutral loci.

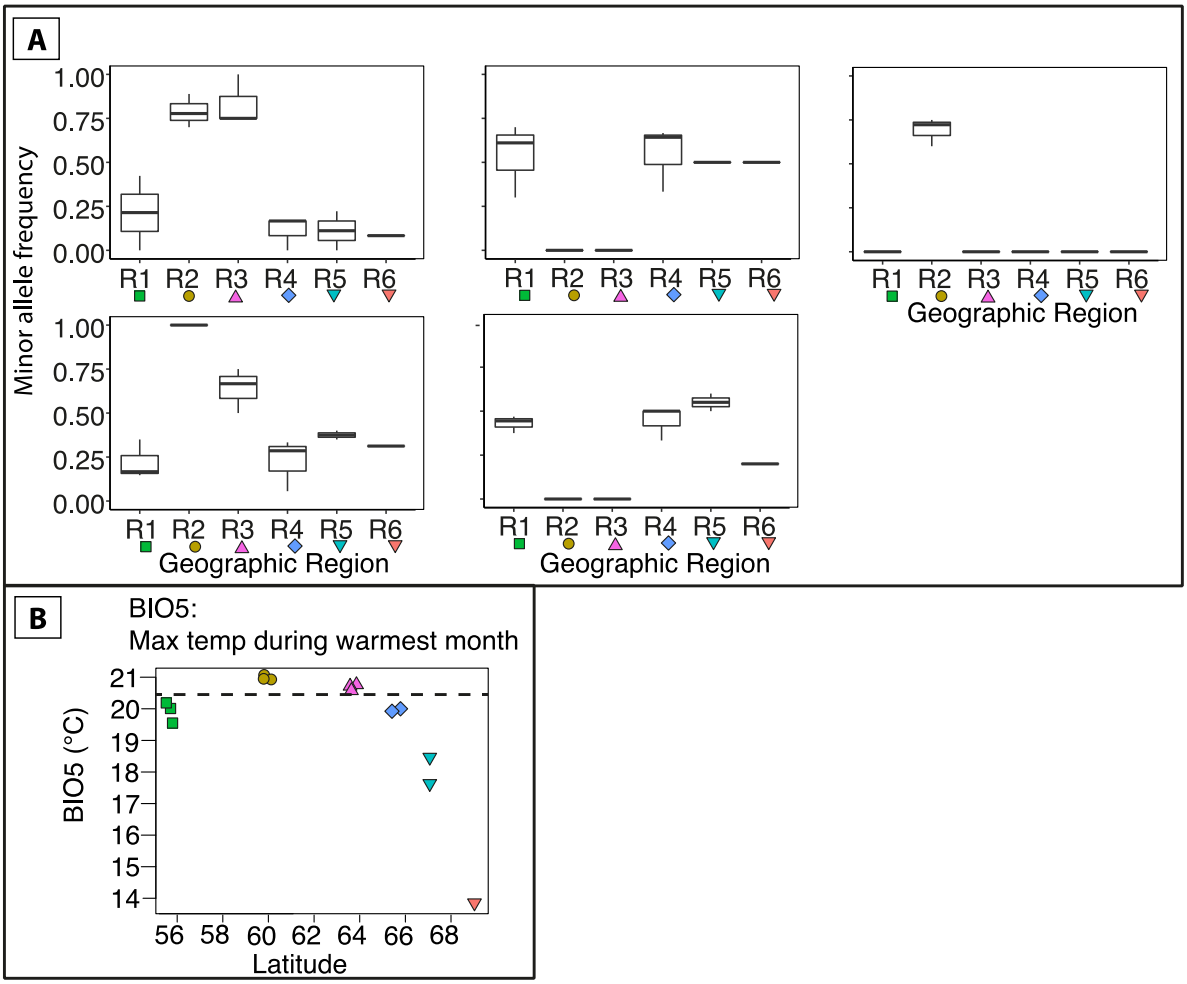


Figure 5.7 Standardised change in the minor allele frequency in five loci from the Outlier dataset that were associated with BIO5 is shown across the latitudinal gradient (A). Allele frequencies were all standardised to range between 0 and 1. Allele frequency was found to be dramatically different in R2 and R3 compared to the rest of the gradient. The change in allele frequency between R1 and R2 explains the dramatic difference between the predicted allele frequencies in the Reference and Outlier datasets (Fig. 5.9). The geographic distribution of BIO5 (B) is shown to illustrate the small change in temperature between R1 and R2. The dashed line indicates the temperature associated with this threshold-like response in the change in allele frequency.

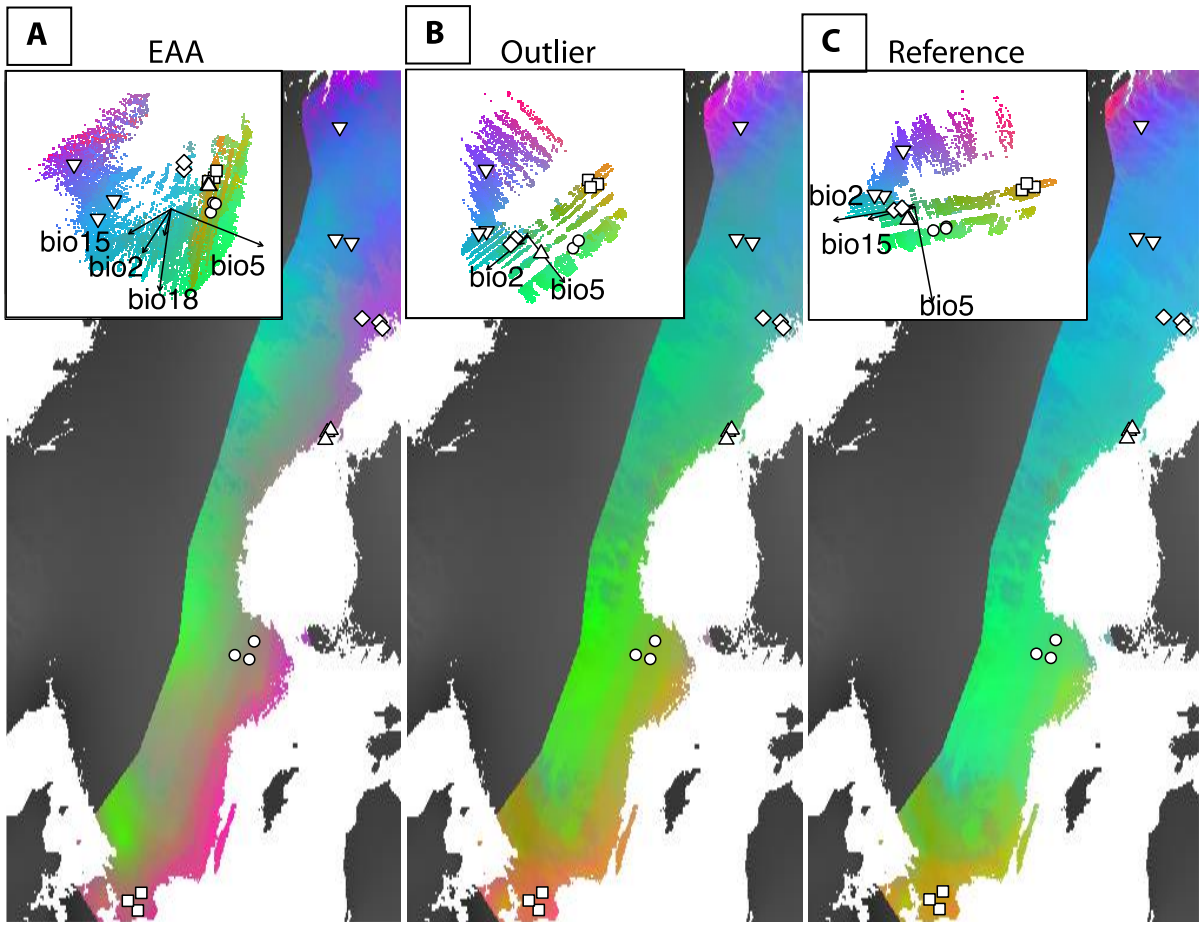


Figure 5.8 Predicted spatial distribution of genomic composition as determined for the EAA (A), Outlier (B), and Reference (C) datasets. The five BioClim variables are transformed by their relative importance in predicting genomic turnover in each dataset, and visualised as a PCA with a colour assigned to the first three principal components. Population genomic composition is expected to be similar on the same colours. The inset in each panel shows a PCA of the transformed BioClim variables, with the most important variables (see Fig. 5.5) shown with arrows. Sampled populations are shown on the PCA and map using the same white symbols as in Fig. 5.1.

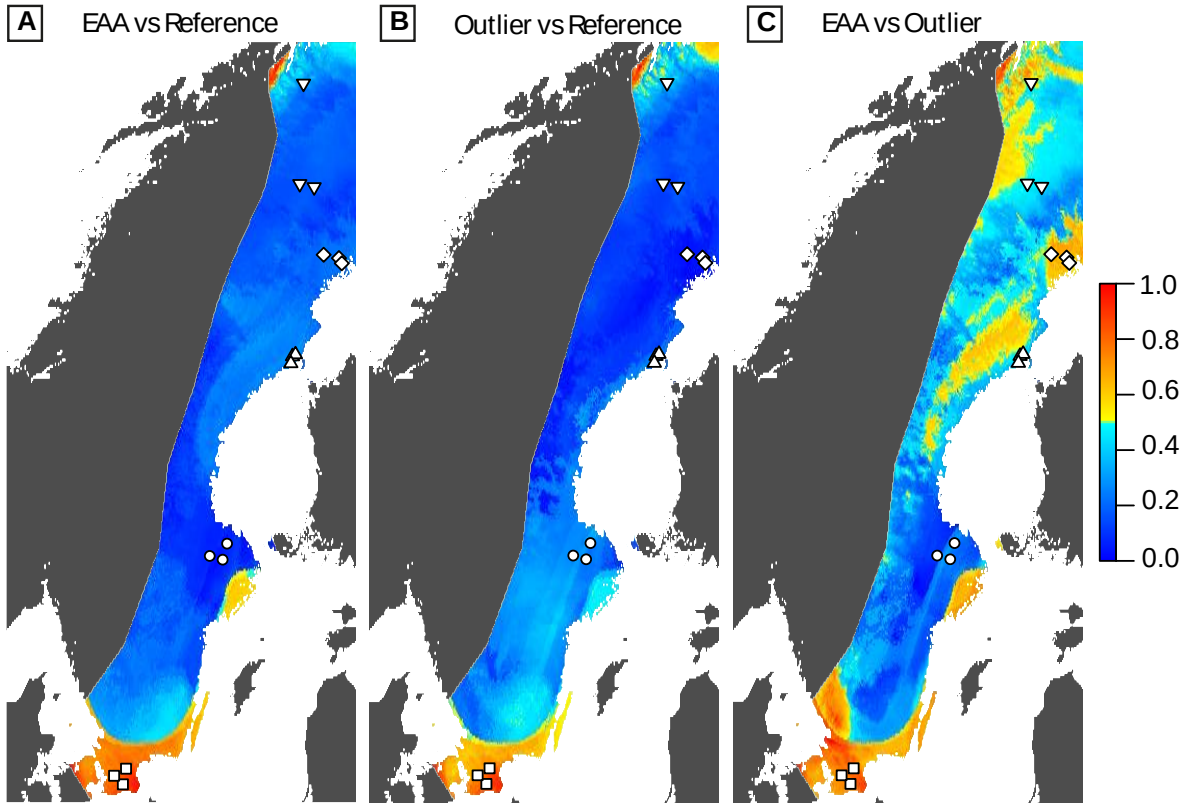


Figure 5.9 Difference in genomic turnover between the reference and adaptive datasets (a & b), and the two adaptive datasets (c). Distances were calculated as the difference between Procrustes residuals in the matrix comparisons and scaled by the maximum distance found for each comparison. Large differences between datasets are indicated with warmer colours. Sampled populations are shown using white symbols (see Fig. 5.1).

Discussion

By combining landscape genetics and genomic turnover analyses we describe how climate and geography structure *R. temporaria* genomic variation across the Scandinavian latitudinal gradient. Genomic variation in this system is strongly related to geography, but also shows evidence of adaptation to climate. A surprisingly large portion of the total genomic variation is attributable to climate variables (38%) or geographically structured climate (30%). We also find a threshold response in to BIO5 (maximum temperature in the warmest month), implying that an thermal threshold occurs in southern Sweden. Finally, our results show that the biggest mismatch between neutral and adaptive allele frequencies occurs in southern Sweden, largely driven by the threhold response to BIO5. Our results show that an analysis of the geographic distribution of genomic variation in *R. temporaria* provide important insights into the climatic drivers and potential adaptive thresholds across a well-studies system.

Population Structure

We found strong population structure across the latitudinal gradient, with the biggest divergence in allele frequencies between R1 and the rest of Sweden. Sequential pairing off of the rest of the populations and strong signals of IBD are indicative of an expansion from a southern colonisation. These results support previous work based on mtDNA that has suggested a single colonisation route and northward expansion in Scandinavia by *R. temporaria* (Palo *et al.* 2004). Previous phylogenetic work from the same study indicated that all populations in Scandinavia and northwards assign to the eastern mitochondrial haplogroup (Palo *et al.* 2004), and the contact zone between the eastern and western lineages has been described in Northern Germany (Schmeller *et al.* 2008). However, population assignment analysis based on microsatellite data found that some southern Scandinavian populations (including from R1) assigned up to 100% to the Western lineage based on multilocus genotypes (Palo *et al.* 2004). This explains the divergence of the southern populations, perhaps even up to R2, from the rest of the gradient. Overall our results support previous work that asserted that two mitochondrial lineages

colonised Scandinavia from the south; via Denmark and Sweden. A contact zone between these lineages in southern Sweden resulted in the divergent allele frequencies in this region compared to the rest of the gradient. The eastern mitochondrial haplotype occur throughout extant Scandinavian populations, which suggests that gene flow from the populations with the eastern haplogroup have since swamped and replaced the western haplogroup (Palo *et al.* 2004).

Geography and Climate determine genotype

We found that climate and geographically structured climate explained a large proportion (76.6%) of the total genomic variation. Climate independent of geography and geographically structured climate explained similar amounts of variation (38% and 30%, respectively). Strong clinal genetic structure across the Scandinavian latitudinal gradient which has been attributed to consistent selection gradients co-varying with geography (Palo *et al.* 2003b, 2004; Cano *et al.* 2004). However, our results suggest that a significant proportion of the adaptive divergence across the gradient could be associated with climate variables that are not latitudinally ordered. Season length (number of days above 6 °C, the developmental threshold for *R. temporaria* tadpoles; Laurila *et al.* 2001; Laugen *et al.* 2003b; Muir *et al.* 2014a) and temperature during larval development (30 days after spawning) are two environmental variables that are commonly attributed to the latitudinal adaptive divergence (e.g. Laugen *et al.* 2003b, 2005a). While season length is latitudinally ordered, water temperatures during the larval phase peak at mid-latitudes (Laugen *et al.* 2003b). Common garden experiments have found that several larval traits - egg development time, size at hatching, larval growth rate, size at metamorphosis, and resting metabolic rate - follow this curvilinear distribution across latitude (Pahkala *et al.* 2002; Laugen *et al.* 2003b, 2005b; Palo *et al.* 2003b; Lindgren & Laurila 2005, 2009). Adult body size, skeletal growth, and lifetime activity follow the same curvilinear distribution, and are maximised at mid-latitudes (Laugen *et al.* 2005b; Hjernquist *et al.* 2012).

Together with our results show support for latitudinally ordered adaptive divergence, but also present evidence that climatic drivers that are not latitudinally ordered are important for adaptation.

Non-linear changes in allele frequencies and threshold effects

We find a strong threshold effect in a subset of Outlier loci associated with BIO5 (maximum temperature during the warmest month). This suggests that there is a physiological threshold in response to BIO5 (or something related to BIO5). Thesholds in polygenic traits are likely to be common in heterogenous environments (Roff 1996). Indeed, we find that four of the five BioClim variables showed an elevated cumulative importance for the adaptive loci compared with the Reference datasets. Of these, BIO2 (mean diurnal range in temperature) and BIO18 (precipitation during the warmest quarter) show evidence of thresholds at which allele frequencies change more rapidly than in the Reference dataset. Geographically the BIO5 threshold separates R2 and R3 from the rest of the gradient. We found dramatically different allele frequencies in a set of Outlier loci associated with BIO5 in this region. Comparison in genomic turnover between the datasets identified the transition between R1 and R2 to diverge the most between the Reference and adaptive datasets. This is indicative of strong adaptive divergence in this region.

Adaptive divergence across the latitudinal gradient in Europe has been extensively documented in plants and animals (e.g. Laugen *et al.* 2003b; Debieu *et al.* 2013; Vergeer & Kunin 2013). Much of this work is based on common garden and reciprocal transplant experiments have established divergence in various phenotypes across the environmental gradient. Landscape and population genomic methods provide a powerful approach to complement and extend these results in several ways. These include determining the proportion of genomic variation associated with adaptation, identifying the genomic underpinnings and architecture of adaptation, identifying important climatic drivers of adaptive divergence, and identifying adaptive thresholds in response to a specific variable. More generally, this approach can have valuable conservation implications, particularly for mitigating the loss of biodiversity due to climate change. One of the most valuable outcomes lies in identifying populations where a small change in environment will result in a large mismatch between genotype and environment. These populations are particularly vulnerable to extinction, and conservation management action would have to be carefully considered.

References

Akaike H (1973) Information theory and an extension of the maximum likelihood principle. In: *2nd International Symposium on Information Theory, Tsahkadsor, Armenia, USSR*, pp. 267–281. Budapest:Akadémiai Kiadó.

Alberto F, Niort J, Derory J *et al.* (2010) Population differentiation of sessile oak at the altitudinal front of migration in the French Pyrenees. *Molecular Ecology*, **19**, 2626–2639.

Alton LA, Franklin CE (2017) Drivers of amphibian declines: effects of ultraviolet radiation and interactions with other environmental factors. *Climate Change Responses*, **4**, DOI 10.1186/s40665-017-0034-7.

Altschup SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic Local Alignment Search Tool. *Journal of Molecular Biology*, **205**, 403–410.

Alvarez N, Thiel-Egenter C, Tribsch A *et al.* (2009) History or ecology? Substrate type as a major driver of spatial genetic structure in Alpine plants. *Ecology Letters*, **12**, 632–640.

Ammann B, Birks HJB, Brooks SJ *et al.* (2000) Quantification of biotic responses to rapid climatic changes around the Younger Dryas — a synthesis. *Paleogeography, Paleoclimatology, Paleoecology*, **159**, 313–347.

Avise JC, Riddle B (2009) Phylogeography : Retrospect and Prospect. *Journal of Biogeography*, **36**, 3–15.

Bachmann JC (2017) Adaptive Divergence across an Elevational Gradient in the Common Frog (Rana temporaria). University of Zurich.

Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, **30**, 2114–2120.

Bonin A (2008) Population genomics: a new generation of genome scans to bridge the gap with functional genomics. *Molecular ecology*, **17**, 3583–4.

Bonin A, Taberlet P, Miaud C, Pompanon F (2006) Explorative genome scan to detect candidate loci for adaptation along a gradient of altitude in the common frog (Rana temporaria). *Molecular biology and evolution*, **23**, 773–83.

Borcard D, Legendre P (2002) All-scale spatial analysis of ecological data by means of principal coordinates of neighbour matrices. *Ecological Modelling*, **153**, 51–68.

Braaker S, Heckel G (2009) Transalpine colonisation and partial phylogeographic erosion by dispersal in the common vole (Microtus arvalis). *Molecular Ecology*, **18**, 2528–2531.

Brady LD, Griffiths RA (2000) Developmental responses to pond desiccation in tadpoles of the British anuran amphibians (Bufo bufo, B. calamita and Rana temporaria). *Journal of Zoology, London*, **252**, 61–69.

Van Buskirk J (2012) Permeability of the landscape matrix between amphibian breeding sites. *Ecology and evolution*, **2**, 3160–7.

Canestrelli D, Cimmaruta R, Costantini V, Nascetti G (2006) Genetic diversity and phylogeography of the Apennine yellow-bellied toad Bombina pachypus, with implications for conservation. *Molecular Ecology*, **15**, 3741–3754.

Canestrelli D, Cimmaruta R, Nascetti G (2008) Population genetic structure and diversity of the Apennine endemic stream frog, Rana italica – insights on the Pleistocene evolutionary history of the Italian peninsular biota. *Molecular Ecology*, **17**, 3856–3872.

Canestrelli D, Nascetti G (2008) Phylogeography of the pool frog Rana (Pelophylax) lessonae in the Italian peninsula and Sicily: Multiple refugia, glacial expansions and nuclear-mitochondrial discordance. *Journal of Biogeography*, **35**, 1923–1936.

Cano JM, Laurila A, Palo J, Merilä J (2004) Population differentiation in G matrix structure due to natural selection in Rana temporaria. *Evolution*, **58**, 2013–2020.

Catchen JM, Amores A, Hohenlohe P, Cresko W, Postlethwait JH (2011) Stacks: Building and Genotyping Loci De Novo From Short-Read Sequences. *G3:Genes, Genomes, Genetics*, **1**, 171–182.

Caye K, Deist TM, Martins H, Michel O, François O (2016) TESS3: Fast inference of spatial population structure and genome scans for selection. *Molecular Ecology Resources*, **16**, 540–548.

CH2014–Impacts (2014) *Toward quantitative scenarios of climate change Impacts in Switzerland*. OCCR, FOEN, MeteoSwiss, C2SM, Agroscope, and ProClim, Bern, Switzerland.

Clement M, Posada D, Crandall KA (2000) TCS : a computer program to estimate gene genealogies. *Molecular Ecology*, **9**, 1657–1659.

Cornetti L, Lemoine M, Hilfiker D *et al.* (2016) Higher genetic diversity on mountain tops: the role of historical and contemporary processes in shaping genetic variation in the bank vole. *Biological Journal of the Linnean Society*, **118**, 233–244.

Croteau MC, Davidson MA, Lean DRS, Trudeau VL (2008) Global Increases in Ultraviolet B Radiation : Potential Impacts on Amphibian Development and Metamorphosis. *Physiological and Biochemical Zoology*, **81**, 743–761.

Crottini A, Andreone F, Kosuch J *et al.* (2007) Fossorial but widespread: the phylogeography of the common spadefoot toad (Pelobates fuscus), and the role of the Po Valley as a major source of genetic variability. *Molecular Ecology*, **16**, 2734–2754.

Danecek P, Auton A, Abecasis G *et al.* (2011) The variant call format and VCFtools. *Bioinformatics*, **27**, 2156–2158.

Dansgaard W, Johnsen SJ, Clausen HB *et al.* (1993) Evidence for general instability of past climate from a 250-kyr ice-core record. *Nature*, **364**, 218–220.

Darnault R, Rolland Y, Braucher R *et al.* (2011) Timing of the last deglaciation revealed by receding glaciers at the Alpine-scale : impact on mountain geomorphology. *Quaternary Science Reviews*, 1–16.

Darriba D, Taboada GL, Doallo R, Posada D (2012) jModelTest 2: more models, new heuristics and parallel computing. *Nature Methods*, **9**, 772.

Debieu M, Tang C, Stich B *et al.* (2013) Co-Variation between seed dormancy, growth rate and flowering time changes with latitude in Arabidopsis thaliana. *PloS one*, **8**, 1–12.

Demesure B, Comps B, Petit RJ (1996) Chloroplast DNA Phylogeography of the Common Beech (Fagus sylvatica L.) in Europe. *Evolution*, **50**, 2515–2520.

Drummond AJ, Suchard MA, Xie D, Rambaut A (2012) Bayesian Phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution*, **29**, 1969–1973.

Dudaniec RY, Spear SF, Richardson JS, Storfer A (2012) Current and historical drivers of landscape genetic structure differ in core and peripheral salamander populations. *PloS one*, **7**.

Eaton DAR (2014) PyRAD: Assembly of de novo RADseq loci for phylogenetic analyses. *Bioinformatics*, **30**, 1844–1849.

Ellis N, Smith SJ, Pitcher CR (2012) Gradient forests : calculating importance gradients on physical predictors. *Ecology*, **93**, 156–168.

Excoffier L, Lischer HEL (2010) An Integrated Software Package for Population Genetics Data Analysis. *Molecular Ecology Resources*, **10**, 564–567.

Fitzpatrick SW, Gerberich JC, Kronenberger JA, Angeloni LM, Funk WC (2015) Locally adapted traits maintained in the face of high gene flow. *Ecology Letters*, **18**, 37–47.

Fitzpatrick MC, Keller SR (2015) Ecological genomics meets community-level modelling of biodiversity: mapping the genomic landscape of current and future environmental adaptation. *Ecology Letters*, **18**, 1–16.

Forester BR, Jones MR, Joost S, Landguth EL, Lasky JR (2016) Detecting spatial genetic signatures of local adaptation in heterogeneous landscapes. *Molecular Ecology*, **25**, 104–120.

Fox J, Weisberg S (2011) *An {R} Companion to Applied Regression*. Thousand Oaks CA: Sage.

Freedman AH, Thomassen HA, Buermann W, Smith TB (2010) Genomic signals of diversification along ecological gradients in a tropical lizard. *Molecular ecology*, **19**, 3773–3788.

Frichot E, François O (2015) LEA : An R package for landscape and ecological association studies (B O'Meara, Ed,). *Methods in Ecology and Evolution*, **6**, 925–929.

Frichot E, Schoville SD, Bouchard G, François O (2013) Testing for associations between loci and environmental gradients using latent factor mixed models. *Molecular Biology and Evolution*, **30**, 1687–99.

Frichot E, Schoville S, de Villemereuil P, Gaggiotti OE, François O (2015) Detecting adaptive evolution based on association with ecological gradients: Orientation matters! *Heredity*, **115**, 22–28.

Garcia VOS, Ivy C, Fu J (2017) Syntopic frogs reveal different patterns of interaction with the landscape: A comparative landscape genetic study of Pelophylax nigromaculatus and Fejervarya limnocharis from central China. *Ecology and Evolution*, **7**, 9294–9306.

Gosner KL (1960) A Simplified Table for Staging Anuran Embryos and Larvae with Notes on Identification. *Herpetologica*, **16**, 183–190.

Goudet J (2005) Hierfstat, a package for R to compute and test hierarchical F-statistics. *Molecular Ecology Notes*, **5**, 184–186.

Gugerli F, Englisch T, Niklfeld H *et al.* (2008) Relationships among levels of biodiversity and the relevance of intraspecific diversity in conservation – a project synopsis. *Perspectives in Plant Ecology, Evolution and Systematics*, **10**, 259–281.

Günther T, Coop G (2013) Robust identification of local adaptation from allele frequencies. *Genetics*, **195**, 205–220.

Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series*, **41**, 95–98.

Harrisson KA, Amish SJ, Pavlova A *et al.* (2017) Signatures of polygenic adaptation

associated with climate across the range of a threatened fish species with high genetic connectivity. *Molecular ecology*, **26**, 6253–6269.

Hecht BC, Matala AP, Hess JE, Narum SR (2015) Environmental adaptation in Chinook salmon (Oncorhynchus tshawytscha) throughout their North American range. *Molecular Ecology*, **24**, 5573–5595.

Heiri C, Bugmann H, Tinner W, Heiri O, Lischke H (2006) A model-based reconstruction of Holocene treeline dynamics in the Central Swiss Alps. *Journal of Ecology*, **94**, 206–216.

Hermisson J, Pennings PS (2005) Soft sweeps: molecular population genetics of adaptation from standing genetic variation. *Genetics*, **169**, 2335–52.

Hewitt GM (1996) Some genetic consequences of ice ages, and their role in divergence and speciation. *Biological Journal of the Linnean Society*, **58**, 247–276.

Hewitt G (1999) Post-glacial re-colonization of European biota. *Biological Journal of the Linnean Society*, **68**, 87–112.

Hewitt G (2000) The genetic legacy of the Quaternary ice ages. *Nature*, **405**, 907–913.

Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A (2005) Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, **25**, 1965–1978.

Hitchings SP, Beebee TJC (1997) Genetic substructuring as a result of barriers to gene flow in urban Rana temporaria (common frog) populations: implications for biodiversity conservation. *Heredity*, **79**, 117–127.

Hjernquist MB, Soderman F, Jonsson KI *et al.* (2012) Seasonality determines patterns of growth and age structure over a geographic gradient in an ectothermic vertebrate. *Oecologia*, **170**, 641–649.

Hoban S, Kelley JL, Lotterhos KE *et al.* (2016) Finding the Genomic Basis of Local Adaptation: Pitfalls, Practical Solutions, and Future Directions. *The American Naturalist*, **188**, 000–000.

Hoffmann A a, Sgrò CM (2011) Climate change and evolutionary adaptation. *Nature*, **470**, 479–85.

Huelsenbeck JP, Ronquist F (2001) MRBAYES : Bayesian inference of phylogenetic trees. *Bioinformatics Application Note*, **17**, 754–755.

Ilut DC, Nydam ML, Hare MP (2014) Defining loci in restriction-based reduced representation genomic data from nonmodel species: Sources of bias and

diagnostics for optimal clustering. *BioMed Research International*, **2014**, 9 pages.

Jombart T (2008) Adegenet: A R package for the multivariate analysis of genetic markers. *Bioinformatics*, **24**, 1403–1405.

Jombart T, Ahmed I (2011) adegenet 1.3-1: New tools for the analysis of genome-wide SNP data. *Bioinformatics*, **27**, 3070–3071.

Jones MR, Forester BR, Teufel AI *et al.* (2013) Integrating landscape genomics and spatially explicit approaches to detect loci under selection in clinal populations. *Evolution*, **67**, 3455–68.

Kawecki TJ, Ebert D (2004) Conceptual issues in local adaptation. *Ecology Letters*, **7**, 1225–1241.

Lasky JR, Des Marais DL, McKay J *et al.* (2012) Characterizing genomic variation of Arabidopsis thaliana: the roles of geography and climate. *Molecular Ecology*, **21**, 5512–5529.

Laugen AT, Kruuk LEB, Laurila A *et al.* (2005a) Quantitative genetics of larval life-history traits in Rana temporaria in different environmental conditions. *Genetics Research*, **86**, 161–170.

Laugen AT, Laurila A, Jönsson KI, Söderman F, Merilä J (2005b) Do common frogs (*Rana temporaria*) follow Bergmann's rule? *Evolutionary Ecology Research*, **7**, 717–731.

Laugen AT, Laurila A, Merilä J (2002) Maternal and genetic contributions to geographical variation in Rana temporaria larval life-history traits. *Biological Journal fo the Linnean Society*, **76**, 61–70.

Laugen AT, Laurila A, Merilä J (2003a) Latitudinal and temperature-dependent variation in embryonic development and growth in Rana temporaria. *Oecologia*, **135**, 548–554.

Laugen AT, Laurila A, Rasanen K, Merilä J (2003b) Latitudinal countergradient variation in the common frog (Rana temporaria) development rates – evidence for local adaptation. *Journal of Evolutionary Biology*, **16**, 996–1005.

Laurila A, Karttunen S, Meril J (2002) Adaptive Phenotypic Plasticity and Genetics of Larval Life Histories in Two Rana Temporaria Populations. *Evolution*, **56**, 617–627.

Laurila A, Pakkasmaa S, Merila J (2001) Influence of Seasonal Time Constraints on Growth and Development of Common Frog Tadpoles: A Photoperiod Experiment. *Oikos*, **95**, 451–460.

Legendre P, Fortin M-J (2010) Comparison of the Mantel test and alternative approaches for detecting complex multivariate relationships in the spatial analysis of genetic data. *Molecular Ecology Resources*, **10**, 831–844.

Legendre P, Legendre LF (2012) *Numerical ecology*. Elsevier.

Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics Application Note*, **25**, 1451–1452.

Lindgren B, Laurila A (2005) Proximate causes of adaptive growth rates: growth efficiency variation among latitudinal populations of Rana temporaria. *Journal of Evolutionary Biology*, **18**, 820–828.

Lindgren B, Laurila A (2009) Physiological variation along a geographical gradient: is growth rate correlated with routine metabolic rate in Rana temporaria tadpoles? *Biological Journal of the Linnean Society*, **98**, 217–224.

Loman J, Claesson D (2003) Plastic response to pond drying in tadpoles Rana temporaria: tests of cost models. *Evolutionary Ecology Research*, **5**, 179–194.

Lugon-Moulin N, Hausser J (2002) Phylogeographical structure, postglacial recolonization and barriers to gene flow in the distinctive Valais chromosome race of the common shrew (Sorex araneus). *Molecular Ecology*, **11**, 785–794.

Luikart G, England PR, Tallmon D, Jordan S, Taberlet P (2003) The power and promise of population genomics: from genotyping to genome typing. *Nature Reviews Genetics*, **4**, 981–94.

Luu K, Bazin E, Blum MGB (2016) pcadapt: An R package to perform genome scans for selection based on principal component analysis. *Molecular Ecology Resources*, **33**, 67–77.

Manel S, Perrier C, Pratlong M *et al.* (2016) Genomic resources and their influence on the detection of the signal of positive selection in genome scans. *Molecular ecology*, **25**, 170–184.

Manel S, Poncet BN, Legendre P, Gugerli F, Holderegger R (2010) Common factors drive adaptive genetic variation at different spatial scales in Arabis alpina. *Molecular ecology*, **19**, 3824–35.

Marquis O, Miaud C (2008) Variation in UV sensitivity among common frog Rana temporaria populations along an altitudinal gradient. *Zoology*, **111**, 309–317.

Marquis O, Miaud C, Lena J-P (2008) Developmental responses to UV-B radiation in common frog Rana temporaria embryos from along an altitudinal gradient. *Population Ecology*, **50**, 123–130.

Mátyás G, Sperisen C (2001) Chloroplast DNA polymorphisms provide evidence for

postglacial re-colonisation of oaks (Quercus spp.) across the Swiss Alps. *Theoretical and Applied Genetics*, **102**, 12–20.

McCain CM, Colwell RK (2011) Assessing the threat to montane biodiversity from discordant shifts in temperature and precipitation in a changing climate. *Ecology Letters*, **14**, 1236–1245.

Merila J, Laurila A, Laugen AT, Rasanen K, Pahkala M (2000) Plasticity in age and size at metamorphosis in Rana temporaria: comparison of high and low latitude populations. *Ecography*, **23**, 457–465.

Merilä J, Laurila A, Laugen AT, Räsänen K, Pahkala M (2000) Plasticity in Age and Size at Metamorphosis in Rana Temporaria: Comparison of High and Low Latitude Populations. *Ecography*, **23**, 457–465.

Messer PW, Petrov DA (2013) Population genomics of rapid adaptation by soft selective sweeps. *Trends in Ecology and Evolution*, **28**, 659–669.

Miaud C, Guyetant R, Elmberg J (1999) Variations in life-history traits in the common frog Rana temporaria (Amphibia: Anura): a literature review and new data from the French Alps. *Journal of Zoology*, **249**, 61–73.

Miaud C, Merilä J (2001) Local adaptation or environmental induction? Causes of population differentiation in alpine amphibians. *Biota*, **2**, 31–50.

Muir AP, Biek R, Mable BK (2014a) Behavioural and physiological adaptations to low-temperature environments in the common frog, Rana temporaria. *BMC Evolutionary Biology*, **14**, 1471–2148.

Muir A, Piek R, Thomas R, Mable B (2014b) Local adaptation with high gene flow: temperature parameters drive adaptation to altitude in the common frog (Rana temporaria). *Molecular ecology*, **23**, 561–574.

Nychka D, Furrer R, Paige J, Sain S (2015) fields: Tools for spatial data.

Oksanen J, Blanchet FG, Kindt R *et al.* (2015) Vegan: community ecology package. R package vegan, vers. 2.2-1.

Pahkala M, Laurila A, Meril J (2002) Effects of ultraviolet-B radiation on common frog Rana temporaria embryos from along a latitudinal gradient. *Oecologia*, **133**, 458–465.

Pahkala M, Laurila A, Merila J (2000) Ambient Ultraviolet-B radiation reduces hatchling size in the common frog Rana temporaria. *Ecography*, **23**, 531–538.

Pahkala M, Laurila A, Merila J (2001) Carry-over effects of ultraviolet-B radiation on larval tness in Rana temporaria. *Proceedings of the Royal Society of London B: Biological Sciences*, **268**, 1699–1706.

Pahkala M, Merila J, Ots I, Laurila A (2003) Effects of ultraviolet-B radiation on metamorphic traits in the common frog Rana temporaria. *Journal of the Zoological Society of London*, **259**, 57–62.

Palo JU, O'Hara RB, Laugen AT *et al.* (2003a) Latitudinal divergence of common frog (Rana temporaria) life history traits by natural selection: evidence from a comparison of molecular and quantitative genetic data. *Molecular ecology*, **12**, 1963–1978.

Palo JU, O'Hara RB, Laugen AT *et al.* (2003b) Latitudinal divergence of the common frog (Rana temporaria) life history traits by natural selection: evidence from a comparison of molecular quantitative genetic data. *Molecular Ecology*, **12**, 1963–1978.

Palo JU, Schmeller DS, Laurila A *et al.* (2004) High degree of population subdivision in a widespread amphibian. *Molecular Ecology*, **13**, 2631–2644.

Parisod C (2008) Postglacial recolonisation of plants in the western Alps of Switzerland. *Botanica Helvetica*, **118**, 1–12.

Parmesan C (2006) Ecological and Evolutionary Responses to Recent Climate Change. *Annual Review of Ecology, Evolution, and Systematics*, **37**, 637–669.

Parmesan C, Yohe G (2003) A globally coherent fingerprint of climate change impacts across natural systems. *Nature*, **421**, 37–42.

Pennings PS, Hermisson J (2006) Soft sweeps II - Molecular population genetics of adaptation from recurrent mutation or migration. *Molecular Biology and Evolution*, **23**, 1076–1084.

Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE (2012) Double Digest RADseq: An Inexpensive Method for De Novo SNP Discovery and Genotyping in Model and Non-Model Species. *PloS one*, **7**.

Polechová J, Barton NH (2015) Limits to adaptation along environmental gradients. *Proceedings of the National Academy of Sciences of the United States of America*, **112**, 6401–6406.

Poncet BN, Herrmann D, Gugerli F *et al.* (2010) Tracking genes of ecological relevance using a genome scan in two independent regional population samples of Arabis alpina. *Molecular ecology*, **19**, 2896–907.

Pritchard JK, Di Rienzo A (2010) Adaptation - not by sweeps alone. *Nature Reviews Genetics*, **11**, 665–7.

Purcell S, Neale B, Todd-Brown K *et al.* (2007) PLINK: A tool set for whole-genome association and population-based linkage analyses. *American Journal of Human*

Genetics, **81**, 559–575.

Raj A, Stephens M, Pritchard JK (2014) FastSTRUCTURE: Variational inference of population structure in large SNP data sets. *Genetics*, **197**, 573–589.

Rehm EM, Olivas P, Stroud J, Feeley KJ (2015) Losing your edge: climate change and the conservation value of range-edge populations. *Ecology and evolution*, **5**, 4315–4326.

Rellstab C, Fischer MC, Zoller S *et al.* (2016) Local adaptation (mostly) remains local: reassessing environmental associations of climate-related candidate SNPs in Arabidopsis halleri. *Heredity*, 1–9.

Rellstab C, Gugerli F, Eckert AJ, Hancock AM, Holderegger R (2015) A practical guide to environmental association analysis in landscape genomics. *Molecular Ecology*, **24**, 4348–70.

Rigling A, Bigler C, Eilmann B *et al.* (2013) Driving factors of a vegetation shift from Scots pine to pubescent oak in dry Alpine forests. *Global Change Biology*, **19**, 229–240.

Rodrigues N, Betto-Colliard C, Jourdan-Pineau H, Perrin N (2013) Within-population polymorphism of sex-determination systems in the common frog (Rana temporaria). *Journal of evolutionary biology*, **26**, 1569–1577.

Roesti M, Salzburger W, Berner D (2012) Uninformative polymorphisms bias genome scans for signatures of selection. *BMC Evolutionary Biology*, **12**, 94.

Roff DA (1996) The evolution of threshold traits in animals. *The Quarterly Review of Biology*, **71**, 3–35.

Rogers AR, Harpending H (1992) Population Growth Makes Waves in the Distribution of Pairwise Genetic Differences. *Molecular biology and evolution*, **9**, 552–569.

Rogivue A, Graf R, Parisod C, Holderegger R, Gugerli F (2018) The phylogeographic structure of Arabis alpina in the Alps shows consistent patterns across different types of molecular markers and geographic scales. *Alpine Botany*, **0**, 0.

Rousset F (1997) Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. *Genetics*, **145**, 1219–1228.

Roy K, Valentine JW, Jablonski D, Kidwell SM (1996) Scales of climatic variability and time averaging in Pleistocene biotas: implications for ecology and evolution. *Trends in Ecology & Evolution*, **11**, 458–463.

Schmeller DS, Palo JU, Merilä J (2008) A contact zone between two distinct Rana temporaria lineages in northern Germany. *Alytes*, **25**, 93–98.

Schneider S, Excoffier L (1999) Estimation of Past Demographic Parameters From the Distribution of Pairwise Differences When the Mutation Rates Vary Among Sites: Application to Human Mitochondrial DNA. *Genetics*, **152**, 1079–1089.

Schweizer RM, VonHoldt BM, Harrigan R *et al.* (2016) Genetic subdivision and candidate genes under selection in North American grey wolves. *Molecular Ecology*, **25**, 380–402.

Sillero N, Campos J, Bonardi A *et al.* (2014) Updated distribution and biogeography of amphibians and reptiles of Europe. *Amphibia-Reptilia*, **35**, 1–31.

Sork VL, Aitken SN, Dyer RJ *et al.* (2013) Putting the landscape into the genomics of trees: approaches for understanding local adaptation and population responses to changing climate. *Tree Genetics & Genomes*, **9**, 901–911.

Ståhlberg F, Olsson M, Uller T (2001) Population divergence of developmental thermal optima in Swedish common frogs, Rana temporaria. *Journal of Evolutionary Biology*, **14**, 755–762.

Stamatakis A (2006) RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics Application Note*, **22**, 2688–2690.

Stapley J, Reger J, Feulner PGD *et al.* (2010) Adaptation genomics: the next generation. *Trends in ecology & evolution*, **25**, 705–12.

Stefani F, Gentilli A, Sacchi R *et al.* (2012) Refugia within refugia as a key to disentangle the genetic pattern of a highly variable species: The case of Rana temporaria Linnaeus, 1758 (Anura, Ranidae). *Molecular Phylogenetics and Evolution*, **65**, 718–726.

Stehlik I, Blattner FR, Holderegger R, Bachmann K (2002) Nunatak survival of the high Alpine plant Eritrichium nanum (L.) Gaudin in the central Alps during the ice ages. *Molecular Ecology*, **11**, 2027–2036.

Taberlet P, Fumagalli L, Wust-Saucy A-G, Cossons J-F (1998) Comparative phylogeography and postglacial colonization routes in Europe. *Molecular Ecolgy*, **7**, 453–464.

Teacher AGF, Garner TWJ, Nichols RA (2009) European phylogeography of the common frog (Rana temporaria): routes of postglacial colonization into the British Isles, and evidence for an Irish glacial refugium. *Heredity*, **102**, 490–6.

Templeton AR, Crandall KA, Sing CF (1992) Cladistic Analysis of Phenotypic Associations With Haplotypes Inferred From Restriction Endonuclease Mapping and DNA Sequence Data. III. Cladogram Estimation. *Genetics*, **132**, 619–633.

Thomassen H a, Cheviron Z a, Freedman AH *et al.* (2010) Spatial modelling and landscape-level approaches for visualizing intra-specific variation. *Molecular Ecology*, **19**, 3532–48.

Tinner W, Theurillat J (2003) Uppermost limit, extent, and fluctuations of the timberline and treeline ecocline in the Swiss Central Alps during the past 11,500 years. *Arctic, Antarctic, and Alpine Research*, **35**, 158–169.

Veith M, Kosuch J, Vences M (2003) Climatic oscillations triggered post-Messinian speciation of Western Palearctic brown frogs (Amphibia, Ranidae). *Molecular phylogenetics and evolution*, **26**, 310–327.

Veith M, Vences M, Vieites DR, Nieto-roman S, Palanca A (2002) Genetic differentiation and population structure within Spanish common frogs (Rana temporaria complex; Ranidae, Amphibia). *Folia Zoologica*, **51**, 307–318.

Vences M, Hauswaldt JS, Steinfartz S *et al.* (2013) Radically different phylogeographies and patterns of genetic variation in two European brown frogs, genus Rana. *Molecular phylogenetics and evolution*, **68**, 657–70.

Vergeer P, Kunin WE (2013) Adaptation at range margins: common garden trials and the performance of Arabidopsis lyrata across its northwestern European range. *New Phytologist*, **197**, 989–1001.

Vitti JJ, Grossman SR, Sabeti PC (2013) Detecting Natural Selection in Genomic Data. *Annu. Rev. Genet*, **47**, 97–120.

Wang IJ (2012) Environmental and topographic variables shape genetic structure and effective population sizes in the endagered Yosemite toad. *Diversity and Distributions*, **18**, 1033–1041.

Wang IJ, Glor RE, Losos JB (2013) Quantifying the roles of ecology and geography in spatial genetic divergence. *Ecology letters*, **16**, 175–82.

Weir BS, Cockerham CC (1984) Estimating F-Statistics for the Analysis of Population Structure. *Evolution*, **38**, 1358–1370.

Wigginton JE, Cutler DJ, Abecasis GR (2005) A note on exact tests of Hardy-Weinberg equilibrium. *American Journal of Human Genetics*, **76**, 887–93.

Willis KJ, Whittaker RJ (2008) The refugial debate. *Science*, **287**, 1406–1407.

Yannic G, Basset P, Hausser J (2008) Phylogeography and recolonization of the Swiss Alps by the Valais shrew (Sorex antinorii), inferred with autosomal and sex-specific markers. *Molecular Ecology*, **17**, 4118–4133.

Yannic G, Pellissier L, Dubey S *et al.* (2012) Multiple refugia and barriers explain the phylogeography of the Valais shrew, Sorex antinorii (Mammalia: Soricomorpha).

Biological Journal of the Linnean Society, **105**, 864–880.

Yeaman S (2015) Local Adaptation by Alleles of Small Effect. *The American Naturalist*, **186**, S74–S89.

Appendix Chapter 5

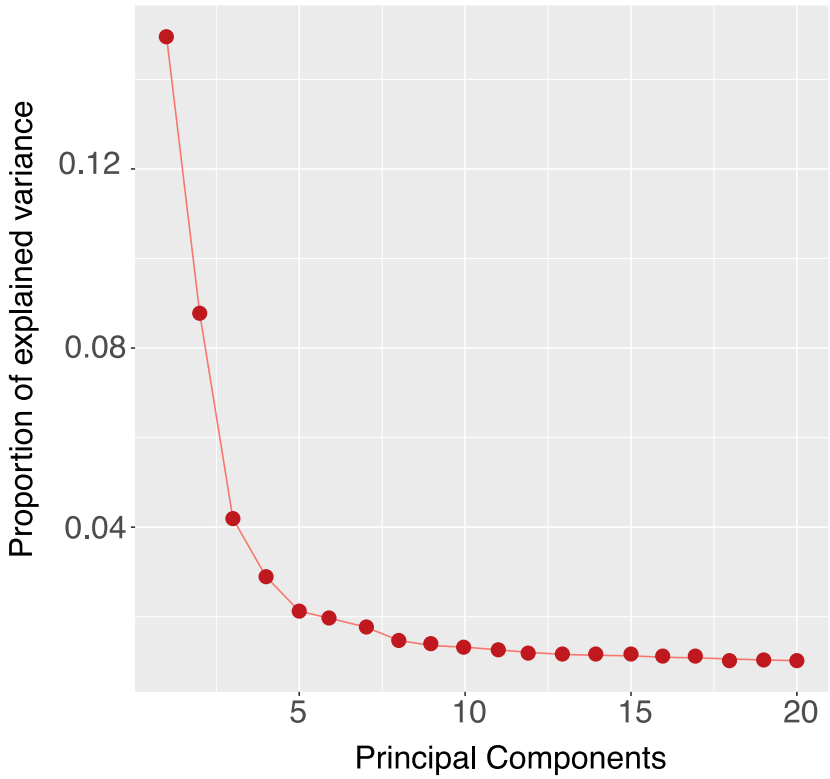


Figure S5.1 Screeplot of the proportion of variance in genomic data explained by the first 20 principal components as calculated using *PCAdapt.* Most of the variance is explained by the first 5 principal components, thus K=5 was chosen.

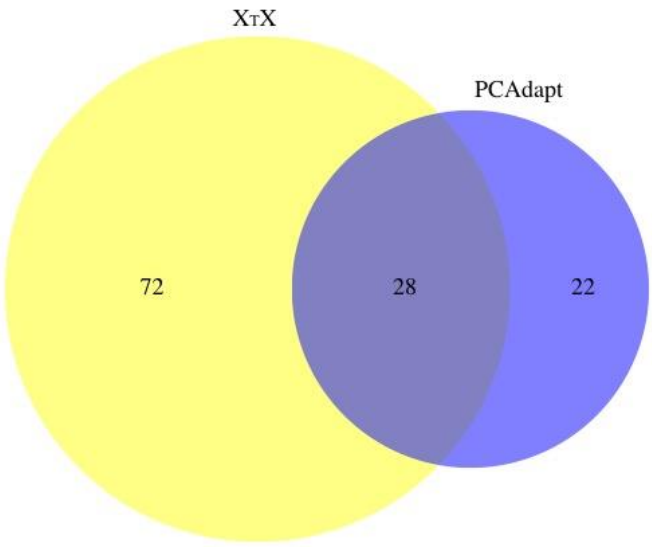


Figure S5.2 Venn Diagram of the outlier loci identified using PCAdapt and the X_TX statistic in BayEnv2. The final dataset was comprised of 122 unique loci.

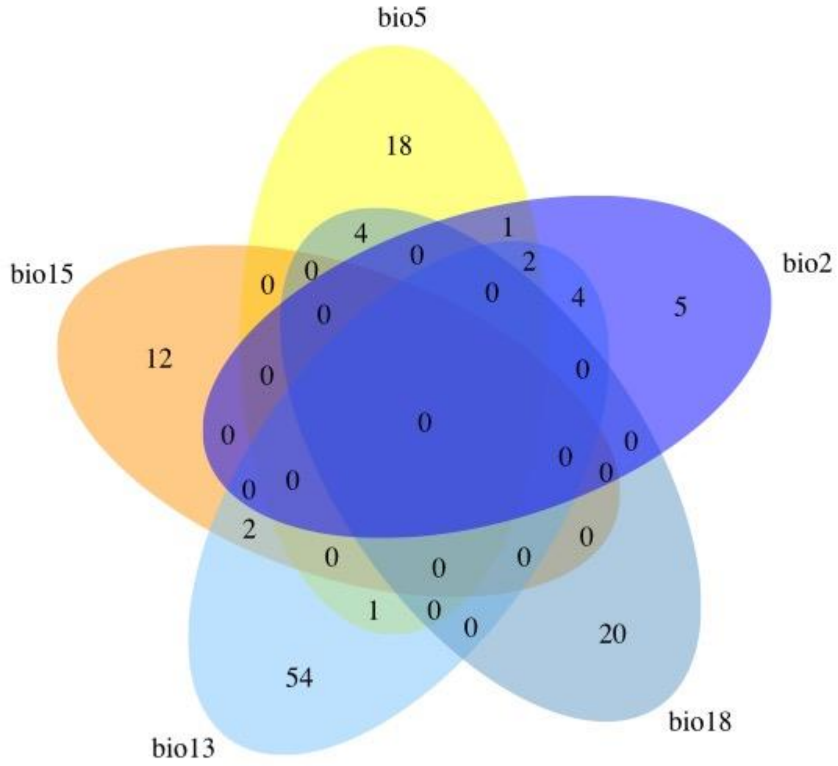


Figure S5.3 Venn Diagram of loci identified by BayEnv2 as associated with the five BioClim variables used in this study. In total, 6% (123) of the tested loci were identified as candidates.

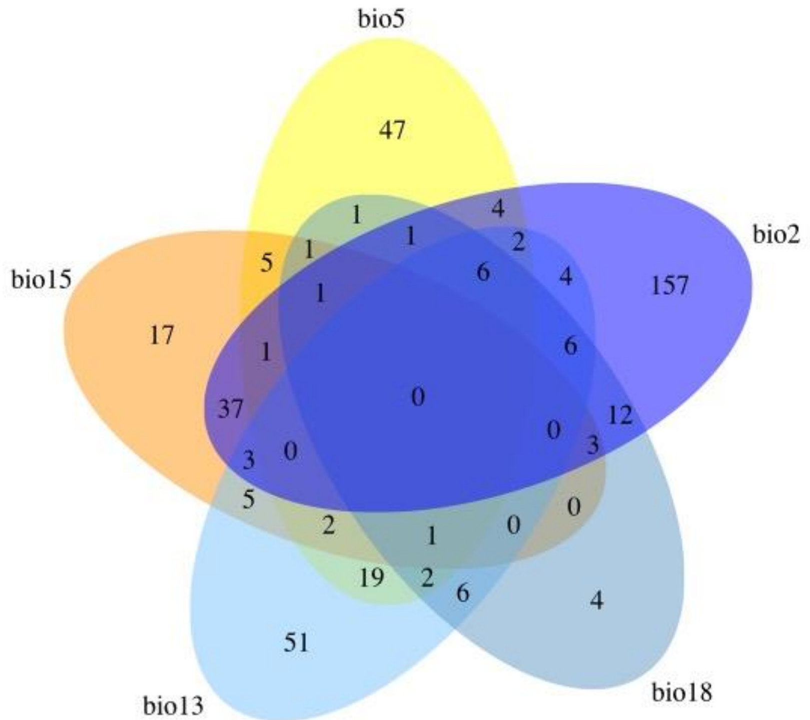


Figure S5.4 Venn Diagram of loci identified by LFMM as associated with the five BioClim variables used in this study. In total 19% (398) of the tested loci were identified as candidates.

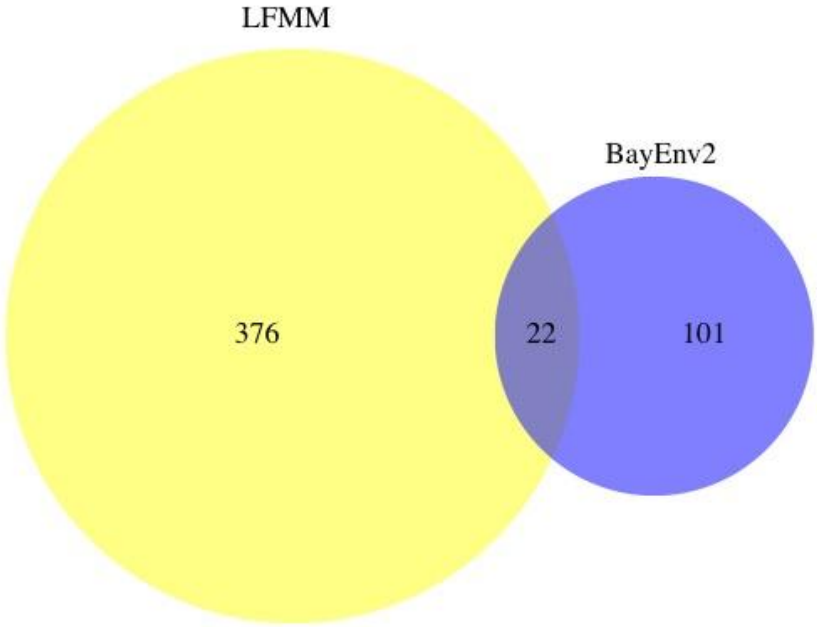


Figure S5.5 Venn diagram of the candidate loci identified by the environmental association analyses. The final dataset comprised 499 unique loci.

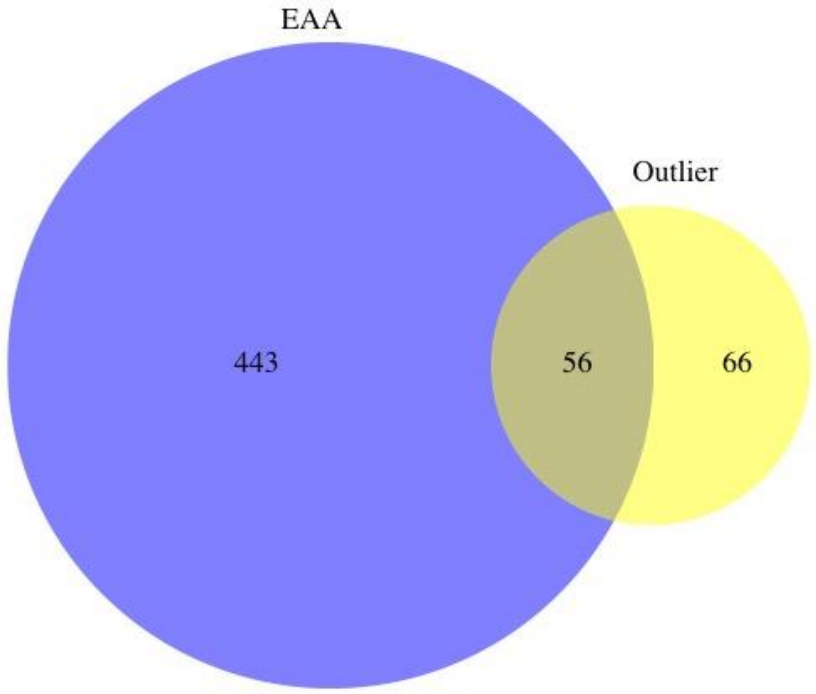


Figure S5.6 Venn diagram of the overlap between the EAA and Outlier datasets. The shared loci represents ~46% of the Outlier dataset, but only ~11% of the EAA dataset.

Table S5.1 Summary of the sequencing data and called loci per population. The number of sequenced individuals (N initial) and the final number of individuals (N final) are shown. RawReads: The number of raw Illumina reads obtained for each population. Passed: the number of reads that passed initial pyRAD. Clusters: the number of loci with a minimum depth of 6 as identified by pyRAD. meanDepth: the mean number of reads per locus with a minimum depth of 6 was calculated across all individuals in the population. Genotyping rate: Ththis was calculated as the mean percentage of sequences per individual that passed all filters and were included in the final dataset. Standard deviation (SD) is shown for all catagories.

Region	Site	N (initial)	RawReads		Passed		Clusters(Depth >5)	
				SD		SD		SD
R1: Skåne ■	Sk_Ho	20	4777179.75	2702878.28	3533151.60	2085059.89	90281.60	50201.83
	Sk_SF	10	4913826.90	1851881.47	3953765.50	1441871.24	110663.50	34024.62
	Sk_SL	19	3810168.42	1040825.93	2773705.63	764790.19	70150.63	18793.34
R2: Uppsala ●	Upp_Gra	10	4416144.00	2344993.07	3573291.10	1902437.60	92547.20	46583.13
	Upp_K	10	5076890.50	1067508.68	4101155.50	839074.08	115332.20	19942.60
	Upp_O	10	3295296.80	1705472.38	2611727.50	1346892.78	70258.50	36344.25
R3: Umeå ▲	Umea_Gr	10	3229892.60	1314863.54	2591338.10	1044612.91	75320.20	31257.61
	Umea_Taf	10	2944441.20	946783.98	2357320.30	767416.29	66456.90	21511.08
	Um.UT3	4	2289299.75	1694447.79	1810977.75	1307251.44	49270.25	33855.17
R4: Luleå ◆	LT1	10	3841540.50	1623871.97	3097627.70	1297984.66	83208.40	33318.78
	LT2	10	1542209.50	1638076.75	1245905.40	1323859.13	33135.10	38098.73
	LT3	10	4027621.30	1266618.78	3233494.70	1028338.04	85031.40	25126.54
R5: Kiruna ▼	Kir_G	10	3355672.60	765677.70	2734730.60	624053.73	79015.90	16884.82
	Kir_L	10	2860095.50	770371.05	2297768.30	635175.85	65786.70	17153.31
R6: Finland ▼	FIN	10	1832168.40	912664.26	1490032.40	912664.26	42623.50	27965.19
Mean		**163**						

Table S5.1 continued

meanDepth (>5)		N (final dataset)	Genotyping rate (%)	
	SD			SD
21.84	2.06	9	90.03	5.26
21.80	2.17	10	95.56	1.47
21.86	1.40	17	92.58	5.02
25.48	5.90	9	95.33	2.01
22.14	1.50	10	95.87	1.24
21.63	2.00	9	93.71	5.30
19.63	1.89	9	95.38	2.31
19.93	2.10	10	92.09	5.39
18.77	2.64	2	72.20	14.37
22.17	2.45	9	97.02	1.21
17.92	2.62	3	97.04	0.77
22.40	1.39	7	96.57	2.16
20.33	1.50	10	95.90	2.30
19.98	1.88	10	94.26	3.28
17.90	1.91	8	86.71	11.46
		132	92.68	

Table S5.2 AMOVA results evaluating the proportion of variation explained by genetic distances at three different levels of organisation. Regions and populations are defined in Table 1 and S1. Significance was assessed by computing the test statistic 1000 times on permutations of each hierarchical level.

Hierarchical Level	Variation (%)	Φ	*P*
Between regions	25.23	52.19	0.00
Variation within regions	6.05	12.51	0.00
Variation within populations	68.72	142.15	0.00

Table S5.3 Pairwise genetic distance (F_{ST}) between all sampled populations.

	Sk.Ho	Sk.SF	Sk.SL	Upp.Gra	Upp.K	Upp.O	Um.Gr	Um.Taf	Um.UT3	LT1	LT2	LT3	Kir.G	Kir.L	FIN
Sk.Ho	-														
Sk.SF	0.04	-													
Sk.SL	0.03	0.04	-												
Upp.Gra	0.15	0.18	0.14	-											
Upp.K	0.13	0.15	0.12	0.09	-										
Upp.O	0.14	0.16	0.12	0.11	0.07	-									
Um.Gr	0.17	0.19	0.16	0.21	0.18	0.19	-								
Um.Taf	0.16	0.17	0.15	0.19	0.17	0.18	0.05	-							
Um.UT3	0.09	0.09	0.06	0.12	0.10	0.11	0.02	0.02	-						
LT1	0.17	0.19	0.16	0.25	0.22	0.22	0.14	0.13	0.06	-					
LT2	0.14	0.14	0.10	0.21	0.17	0.19	0.11	0.10	0.10	0.08	-				
LT3	0.16	0.18	0.14	0.24	0.21	0.22	0.13	0.12	0.07	0.07	0.08	-			
Kir.G	0.15	0.17	0.14	0.20	0.18	0.19	0.12	0.12	0.06	0.11	0.08	0.10	-		
Kir.L	0.16	0.18	0.15	0.22	0.20	0.21	0.14	0.13	0.07	0.11	0.09	0.10	0.07	-	
FIN	0.18	0.20	0.16	0.28	0.25	0.26	0.18	0.17	0.10	0.14	0.13	0.13	0.12	0.12	-